

Blessings of mixed neutrinos

The confirmation that neutrinos have mass and can switch identity is a triumph of careful experiment that opens doors for theoretical physicists. It is not a crisis for existing models, but a route to deeper ones.

Thanks to 1,000 tonnes of exquisitely pure deuterium oxide, two kilometres or so beneath Ontario, we now have hard evidence that the three 'flavours' of neutrino (ν_e , ν_μ and ν_τ) change identity — or oscillate — as they travel from the Sun to the Earth. As John Bahcall describes on page 29, new measurements from the Sudbury Neutrino Observatory, combined with those from the Super-Kamiokande detector in Japan reported in 1998, beautifully resolve a question that has stood unanswered for three decades: whether a deficit in the detected number of solar neutrinos was due to a crisis in the otherwise successful modelling of the Sun and stellar evolution, or to some identity crisis of neutrinos. The solar models are triumphant. Another debate is also resolved: physics tells us that this also means that neutrinos have mass — but, at less than about 10^{-7} of the mass of an electron, not enough to account for all of the mysterious dark matter that pervades the Universe.

But is the result a crisis for neutrino physics, and for high-energy physics more generally? High-energy theorists would have welcomed a crisis any time in the past decade. During that time, the theory that amalgamates models of the strong and weak nuclear forces and those of electromagnetic forces has successfully withstood all experimental challenges. This 'standard model' is one of the great intellectual triumphs of last century: it takes some simple fundamental principles of physical symmetries in space-time, some not-so-simple hypothetical 'internal symmetries' between particle properties, and also the basic postulates of quantum mechanics such as Schrödinger's equation, which links a particle's behaviour, the energy potentials within which it moves and the uncertainty principle. From those, it arrives at a world in which the known fundamental particles arise as excitations of fields, including one, the Higgs field, which gives rise to yet-unseen Higgs particles that interact with known particles to imbue them with mass.

Beyond the standard model

Given its success, whatever comes next, the standard model will be no more fundamentally right or wrong than is Newton's gravitational law, which has been supplanted by general relativity but still provides the basis of successful planetary missions. As with Newton's law, everyone is sure that the standard model is inadequate for explaining more extreme conditions. But, unlike Einstein, nobody has yet convinced everybody that they know just where, theoretically, to go next, although there are many ideas deposited weekly, if not daily, on the Los Alamos preprint server. The critical clues seem more likely to arise not from an unexpectedly falsified prediction of the standard model, but from phenomena discovered in regimes where the standard model is expected to be inadequate.

The fact that neutrinos have mass, therefore, is in no way a crisis for the standard model. Theorists differ about exactly how to represent such masses. Some explanations amount to a small extension of the model; others are more fundamentally different, and are expressed in more broadly predictive frameworks, with a higher degree of mathematical symmetry. These incorporate the standard model at the cost, or benefit, of predicting new particles or new interchangeability

between particles. However, everyone agrees that each of the neutrino flavours observed must be represented as a superposition — a quantum mixture — of underlying basic states, which themselves are the mass-possessing outcomes of interactions with some sort of Higgs field. The standard model is already comfortable with the fact that such mixing can occur. What now needs to be measured or inferred by the Sudbury Neutrino Observatory, and others too, is the nature of the mixing, and the patterns of oscillation not only in empty space, but also, for example, in the interior of the Sun and in other astrophysical environments.

Towards grand unification

Several leading theorists, including Frank Wilczek (see *Nature* **391**, 123–124; 1998) and John Ellis (see <http://xxx.lanl.gov/abs/hep-ph/0008334>), see neutrino masses as one of several pointers to a Grand Unified Theory (GUT) — a seamless unification of the three forces. The nature of the mixing might help to identify just what sort of GUT explicitly describes not only the existence of neutrinos with mass, but also the scale of energy at which the three fundamental forces of nature mentioned above are coupled together with equal strength. Below that energy, according to the standard model, the three forces can be differentially screened by virtual particles and antiparticles that flicker in and out of the vacuum. As Wilczek has pointed out, it must be more than coincidence that the unification energy inferred from neutrino masses, and that derived by theorists from the internal logic of the standard model, are the same, at around 10^{16} GeV. It may even point towards yet higher symmetries, the theories of which (supersymmetry, or SUSY) predict the existence of a plethora of particles that could well be observable using the next generation of particle accelerators.

All neutrinos seen so far are left-handed — their spin, if projected onto their direction of motion, is anticlockwise — whereas the spin of antineutrinos is right-handed. As Ed Witten explains (<http://xxx.lanl.gov/abs/hep-ph/0006332>), considerations of this 'helicity' and of neutrino mass can be related not only to a breakdown of a hypothetical conservation of quantum numbers (lepton and baryon numbers) ascribed to interacting particles, but also to quantum gravity theories that describe black holes and also the unification of gravity with the three other forces at energies of around 10^{19} GeV. Thus, according to theorists such as Witten and Wilczek, a knowledge of the nature of neutrino mixing brings us within theoretical spitting distance of the most elusive goal of all: the theory that unites all the fundamental forces of nature, or the Theory of Everything (TOE).

Particle experimentalists and the rest of us can at last welcome this beginning of an observational engagement with predictions of GUT, SUSY and TOE, via determinations, for example, of mass differences and oscillation probabilities, and inferences of the nature of the basic neutrino states. Whether generated in astronomical objects or in the atmosphere, or fired from near and distant accelerators, or from neutrino 'factories' that are still only gleams in experimentalists' eyes, mixed neutrinos provide us with a new lens through which to glimpse nature's deepest underlying structure. ■





Quotas mooted
Germany strives to entice more women into science posts.
p4



Bomb squad
Health probe for sailors involved in 1950s nuclear tests.
p5



Fallen star
Innovators lose jobs as 'anarchic' research centre closes.
p6



Just the tonic
Hong Kong aims to exploit old-fashioned medicine.
p7

Los Alamos loses physics archive as preprint pioneer heads east

Declan Butler

The Los Alamos preprint server, which has established itself as physicists' favourite place for early circulation of their research, is leaving the New Mexico laboratory to set up shop at Cornell University in New York state.

Paul Ginsparg, who founded the server — now known as arXiv — 10 years ago, is leaving the Los Alamos National Laboratory (LANL) to take up a faculty position at Cornell, and the server will move with him. Cornell plans to expand arXiv's reach into other disciplines, and to use it as a test bed for research into digital libraries.

Ginsparg says growing dissatisfaction with LANL is a major reason for his departure, citing a lack of enthusiasm for the archive among senior staff. Only his former group leader Geoffrey West and library director Rick Luce gave the archive strong support, he says. He adds that the nuclear-weapons laboratory has been shifting its support towards large groups at the expense of individual investigators, and is suffering from declining morale in the wake of recent security scandals.

Los Alamos experienced a painful security clamp-down after Wen Ho Lee, a Taiwanese-born engineer at the laboratory,



Low vantage point: the loss of the preprint server is a blow for the lab on the New Mexico mesa.

was arrested two years ago on espionage charges and then released (see below). The loss of the prestigious server delivers another blow to the laboratory's standing in the scientific community.

William Press, deputy director of the laboratory, says: "We're sorry to see Paul go, but Cornell has created a very unique opportu-

nity for him. We are very proud to have been the incubator of this revolution in scientific publishing." He adds that senior laboratory staff have strongly supported the archive activity, but admits that it was sometimes "a struggle to see where it would fit in" with the laboratory's other activities.

The archive currently receives around

Pentagon offers deal on nuclear weapons book

Irwin Goodwin, Washington

The former director of intelligence at Los Alamos National Laboratory has been told that he can publish his explosive account of China's nuclear-weapons programme — if he accepts heavy censorship of the 500-page manuscript.

Danny Stillman, a physicist who retired from Los Alamos in 1993, filed a suit against the US government on 18 June, arguing that it had unlawfully exceeded its constitutional authority by blocking publication of his manuscript for 18 months.

But on 27 June, Stillman's lawyer, Mark Zaid, was given the manuscript, marked up by Pentagon intelligence officials. "They

released about 85 to 90% of the manuscript and said the rest was highly classified," says Zaid. "I sent the whole thing to Danny to examine. He may end up challenging the omissions on grounds that the information came from the Chinese and doesn't compromise our security in any way."

Stillman made nine trips to China, with government consent, between 1990 and 1999 — six of them as a private citizen after he retired from the lab. His book, *Inside China's Nuclear Weapons Program*, contains what may be the most detailed record ever of the country's nuclear-weapons activities, with information on its 46 nuclear tests and more than 2,000 scientists and engineers

involved in the programme.

The Pentagon has argued that the book contains secret information that, if made public, would damage both US national security and its relations with China.

But some experts think that material in the book would undermine accusations that China relied on spying in the United States to acquire nuclear-weapons expertise. Irving Lerch, the director of international affairs at the American Physical Society, says that releasing Stillman's findings "would go a long way in eliminating the charge that foreign-born scientists, particularly Chinese, can't be trusted with our nuclear secrets" (see News Feature, page 10).



Paul Ginsparg: set to move to Cornell.

\$300,000 in annual funding from the National Science Foundation, the Department of Energy, which runs the lab, and LANL itself.

Ginsparg says that consultation with the archive's advisory board, funding agencies and the American Physical Society, produced a consensus that the operation would enjoy more secure funding and stronger intellectual support at a university than at LANL.

But for Ginsparg, the last straw was his recent salary review, which, he says, described him as "a strictly average performer by overall lab standards; with no particular computer skills contributing to lab programs; easily replaced, and moreover overpaid, according to an external market survey".

LANL officials declined to comment on Ginsparg's case, but said that some recent salary increases at the laboratory have been available only to certain combinations of programmes and individual skills.

Peter Lepage, chair of Cornell's physics department, notes wryly of the LANL assessment: "Evidently their form didn't have a box for: 'completely transformed the nature and reach of scientific information in physics and other fields.'" ■

♦ <http://arXiv.org>

Imported stem cells deepen Germany's ethical divide

Quirin Schiermeier, Munich

The charged debate in Germany over research on human embryonic stem (ES) cells boiled over earlier this week, with the news that at least one scientist has already imported human ES cell lines.

Wolfgang-Michael Franz, a cardiologist at the Ludwig Maximilian University of Munich, revealed on 2 July that in March he received human ES lines, ordered from the WiCell Research Institute in Madison, Wisconsin. He has not yet used the cells, but plans to seek funding from the DFG, Germany's main research granting body, to develop cardiac muscle from the cell lines.

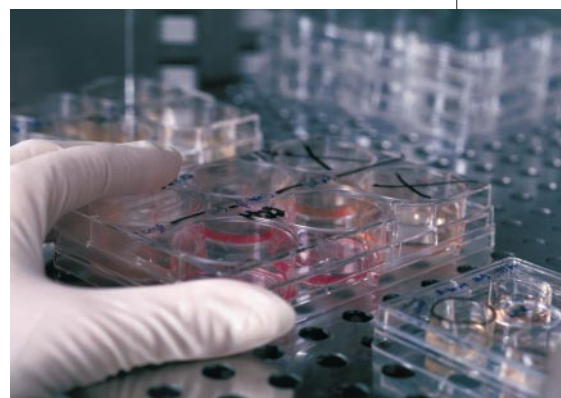
Although German law bans the isolation of human ES cells from 'spare' embryos, their import is not prohibited. But with public feelings running high, the import of cells is intensely controversial. Last week, when it emerged that Stefan Rose-John, a biochemist at the University of Kiel, planned to import stem cells, he was criticized in the media.

Meanwhile, police are assessing the safety of Oliver Brüstle of the University of Bonn, who was the first to apply to the DFG to work on human ES cells. Brüstle received death threats after the *Frankfurter Allgemeine Zeitung* published details of a patent application that included his private address.

The science ministry seems to have been

caught off-guard by the latest revelations. "I call on all researchers planning to use human stem cells to wait for a vote of the national ethics council," says science minister Edelgard Bulmahn. The council is expected to deliver its opinion on whether research should go ahead in the autumn.

The German parliament will vote on a moratorium on the import of stem cells on 5 July. But it is unlikely to find a majority, as the ruling Social Democrat-Green coalition has agreed for now not to impose additional restrictions on biomedical research. ■



Eye of the storm: WiCell's human embryonic stem cells have been imported into Germany.

Quota offered as solution to gender imbalance

Katja Henssel, Munich

Germany is looking at a controversial answer to the perennial problem of gender inequality in the workplace. A law being considered by the German parliament would introduce a quota system for women in government-funded jobs, including those in science.

Currently, only about 5% of top positions

at public research institutes and universities in Germany are held by women, compared with nearly 14% in France and 11% in Italy.

The government wants to change that fast. "By 2005, we want at least 20% of full professorships and other leading scientific positions to be filled with women," says Helga Ebeling, head of the federal science ministry's division for women in education and research.

The legislation, which would cover universities as well as government-funded research institutions, was drafted by the government in the spring and is scheduled to be considered by the parliament in the autumn. It already has the support of the ruling Social Democrat-Green coalition.

German research organizations and prominent scientists say that they welcome action on gender discrimination. But many doubt that a quota system is the best approach to fixing the problem.

"It would certainly cause a quick injection of women into the science system,"

Christiane Nüsslein-Volhard, director of the Max Planck Institute for Developmental Biology in Tübingen, told a meeting on women in science last month at the European Molecular Biology Organization in Heidelberg. "But the idea that you owe your job to a quota has an unpleasant taste," she said. "It devalues your skills."

Klaus Landfried, president of the Association of Universities and Other Higher Education Institutions in Germany, says that a quota system would undermine the authority of women in top scientific positions. "It is the least clever way to improve opportunities for women," he says.

Supporters of a quota system include Sybille Krummacker, a scientist at the national research centre in Jülich (FZJ), who helped to set up Germany's first women-only tenure-track programme. "The programme has encouraged many women to apply," says Krummacker, "and they are not employed just because they are female, but because they are highly qualified." ■



Uneven: only 5% of top positions in German research institutions are held by women.

Europe plots comeback in neutron science

David Adam, London

Researchers are laying plans to wrestle back Europe's leadership in neutron science after powerful neutron-scattering facilities open in the United States and Japan in a few years' time. This summer, the Europeans are putting the finishing touches to plans for a new £1-billion (US\$1.5-billion) neutron facility of their own.

Two of the three key issues underpinning the design of the proposed European Spallation Source (ESS) were resolved at a UK meeting of project planners in Abingdon on 15 June. A final plan will be submitted to European governments and funding agencies by 2003, and ESS supporters hope to obtain permission to start work on the project the following year.

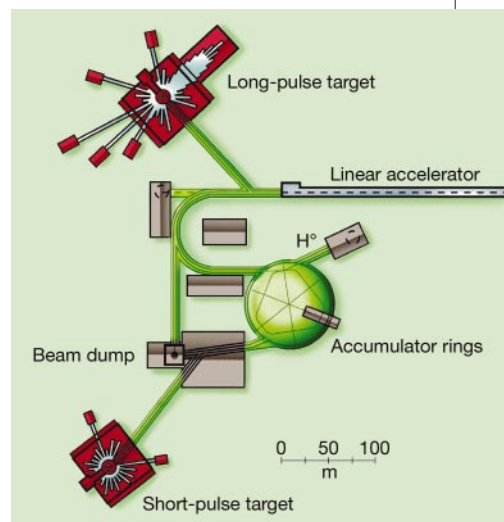
Neutron sources can be used to probe the structure of anything from superconducting materials to proteins, and are therefore an essential tool for physicists, chemists, materials scientists and structural biologists. The interaction of neutrons with light atoms such as hydrogen allows some structures to

be examined more effectively than with other techniques, such as X-ray diffraction.

Scientists planning the ESS say it will be more powerful than both the US Spallation Neutron Source, now under construction at the Oak Ridge National Laboratory in Tennessee, and the planned Japanese spallation neutron source. Europe currently has a clear lead in neutron-scattering science. It is home to the world's two most powerful neutron sources — at the Institut Laue Langevin in Grenoble, France, and ISIS near Oxford in England — and around two-thirds of the discipline's scientific community.

But it could lose some of that lead when the US and Japanese projects, both of which will be more powerful than anything in Europe, open within six years. Even if the ESS is approved in 2004, it will not be operational until at least 2011.

That five-year window will blunt Europe's leadership in the field, predicts Bob Cywinski, a physicist at the University of Leeds who chairs the European Neutron Scattering Association. "We will still be able



Target set-up: the ESS proton stream will be accelerated in the direction of two targets.

to do high-class science, but we will not be at the cutting edge and that is a dangerous scenario," Cywinski says. "And if the ESS does not get built at all then we might as well give up and go home."

If the ESS is constructed, it will be a stand-alone facility dedicated solely to neutron scattering. Spallation sources produce neutrons by first generating a stream of high-energy protons that are crashed into a metal target, chipping off neutrons from the nuclei of the metal atoms. Researchers had suggested that some of the protons produced might be diverted into other projects, such as experimenting with the transmutation of nuclear waste. But the ESS council now says that "strategic considerations" make that option unworkable.

Cywinski is pleased with the decision. "If you have an accelerator that is doing many things, then there is always the suggestion that it might be a compromise and we don't want to compromise the ESS in any way," he says.

The ESS council meeting also decided on the facility's configuration (see figure, above). It has been decided that the proton supply will be directed at two targets — one beam will strike in short proton pulses of 50 hertz and the other in longer, 16.67-hertz pulses.

The third issue, still to be decided, is the specific nature of the proton accelerator that will drive the particles at energies high enough to splinter the metal target atoms.

The site of the ESS is also unresolved, although the existing Forschungszentrum research facility in Jülich, Germany — where the ESS administration is based — is an early favourite. Peter Tindemans, the ESS council chairman, says that interested countries will soon be asked to submit bids.

Gene screens for nuclear veterans

Peter Pockley, Sydney

New Zealand veterans of British nuclear-weapons tests are to be tested for genetic abnormalities. The New Zealand Nuclear Test Veterans' Association said last month that it will conduct a pilot analysis of DNA from 50 veterans.

Around 40,000 troops witnessed or assisted in the 1952–58 tests, which took place in and around Australia and on Christmas Island, and involved 528 New Zealand sailors, with most of the others

coming from Britain and Australia.

Those involved have long complained that the effects have never been fully investigated. The New Zealand veterans claim that their life expectancy has been reduced by at least 10 years, and that there is an unusually high rate of genetic disorders among them and their children.

"Rightly or wrongly, the men who were exposed to the nuclear tests perceive that they now suffer from a range of illnesses that is different from the normal population," says Al Rowland, a cytogeneticist at Massey University in North Palmerston, New Zealand, who will carry out the tests.

Rowland will use stem cells from veterans' bone marrow to look for chromosome damage. In parallel, John Podd, a psychologist from Massey University, will assess the long-term psychological effects of the nuclear trials.

The Australian government also plans to compare the health of the nation's 18,000 test veterans — a full list of whose names was published last month for the first time — with that of the general population.

The New Zealand veterans' association is planning legal action against the British government for allegedly neglecting its duty of care. Previous claims by individuals have been unsuccessful or settled out of court, with no admission of fault by the government.



Testing times: a New Zealand navy frigate sails into the wake of a British H-bomb explosion.

ROYAL NZ NAVY

University reforms fail to quell fears of 'inbreeding'

Xavier Bosch, Barcelona

Responding to mounting evidence that Spain's university system is one of the most inward-looking in the world, the Spanish government has announced sweeping reforms of the system.

But as public comment closed last week on the reform plan, many researchers were complaining that it does not go far enough.

The new measures include arrangements for evaluating and selecting candidates for junior and full professorships; a programme to lure young Spanish scientists back into the country (see *Nature* 410, 1014; 2001); and a new agency to assess university research.

Critics have focused on how candidates will be vetted for faculty positions. The plan would introduce more open, merit-based competition to the early part of the recruitment process, but still leaves the final appointment step in the hands of a panel chosen by the university itself.

Ismael Crespo-Martínez, director of universities at the education ministry, says that the new system will put an end to 'inbreeding' in Spanish universities and will boost the mobility of researchers.

But almost 500 staff from 60 universities have sent an open letter to the education minister, Pilar del Castillo, stating that the "anti-inbreeding measures of the new law are no longer useful if the university panel is not made up of outside professors with no scientific links with the applicant".

One recent survey suggested that the existing system, which gives universities complete autonomy in hiring, leaves Spanish universities ten times more likely to appoint internal candidates to faculty positions than those elsewhere (see *Nature* 410, 14; 2001) ■



Utopian dream in tatters as Starlab crashes to Earth



Jim Giles, Brussels

An ambitious attempt to create a utopian environment for advanced, interdisciplinary research has come to grief. Starlab, a privately funded centre in Brussels that encouraged its researchers to generate innovative ideas, free from the pressure to publish or to develop products, went bust last month. Its 100 staff are now looking for new jobs.

Walter de Brouwer, a former mathematician, established Starlab in 1996 with US\$5 million that he made in computer publishing. Loosely modelled on the Massachusetts Institute of Technology's Media Lab, Starlab started with a core of computer scientists and experts in multimedia technology. As technology stocks soared, de Brouwer raised money from like-minded investors, and expanded Starlab to hire physicists interested in quantum computing and nanotechnology. Biology labs were added in 1999.

Many research centres are trying to forge links between these disciplines. But Starlab differed in its rejection of traditional corporate or academic structures and quality assessment, and its anarchic approach to cross-disciplinary collaboration. De Brouwer's philosophy was that Starlab was created to generate ideas, not products, as reflected in one of the centre's slogans: "We do not make things that work." He describes peer review as "the calibration of mediocrity performed by bearded men".

Starlab provided a haven for scientists with unconventional ideas. Hugo de Garis, for example, was exploring a new approach to artificial intelligence, dubbed 'brain building', in which he tried to simulate animal cognitive abilities by wiring together reconfigurable computer chips. Central to Starlab's ethos was its weekly brainstorming session, known as the 'great ideas club', in which scientists from diverse backgrounds would discuss one another's work and dream up potential projects.

When *Nature* visited Starlab in May, the enthusiasm of many of its scientists was obvious. "People are much less sceptical at



Paradise lost: Starlab's elegant Belgian base will no longer echo to brainstorming sessions.

Starlab," said stem-cell biologist Isabelle Bouhon, who was collaborating with mathematician Keith Still to study the 'emergent properties' of developing cells and tissues.

But by then, the writing was on the wall. Too few of Starlab's 'great ideas' were attracting commercial interest, and investors had declined to buy bonds in the centre. By the end of May, Starlab was unable to pay its staff. And on 11 June, when a group of existing investors withdrew their support, de Brouwer was forced to call in receivers.

Starlab's remaining assets, including its elegant, former embassy building, may have to be sold to cover debts of some US\$3.5 million. Some former staff have been in talks with the Brussels regional government in the hope of resurrecting the centre in a more conventional commercial guise. But for now, things look bleak. "All the scientists have had to be fired," says de Brouwer. "I am trying to ensure that they get five months' salary."

The leaders of a second Starlab, at the Fabra Observatory on the outskirts of Barcelona, hope to survive the demise of its parent by raising further investment on their own. The Spanish lab currently employs 10 researchers and had plans to expand.

As former employees in Brussels ponder their futures, some have concluded that Starlab's experiment in research organization was a step too far. De Garis complains that he was unable to build a research group. "What I really need now is security," he says. ■

♦ <http://www.starlab.org>

Hong Kong seeks secrets of Chinese medicine

David Cyranoski, Hong Kong

Hong Kong is aiming to establish itself as the global centre for the scientific study and commercial exploitation of traditional Chinese medicines.

The former British colony is to found a new Institute of Chinese Medicine, funded with HK\$500 million (US\$64 million) from the charitable trust of the Hong Kong Jockey Club, which runs part of the territory's lucrative gambling industry.

Hong Kong is currently trying to develop a licensing system for suppliers of traditional Chinese remedies. Advocates of the new institute say that it will help suppliers of the remedies to meet new regulatory requirements. At the moment, compounds used in Chinese medicine are not subject to normal regulation in many parts of the world. In the United States, for example, almost anything can be sold as a 'traditional medicine'.

Some scientists believe that a regulation drive will bring necessary legitimacy to the traditional remedies. But others fear that it could confer false legitimacy on compounds that have no proven medical value.

A commission led by Chang-lin Tien, former chancellor of the University of California, Berkeley, has advised the Hong Kong government that traditional Chinese medicine could represent a promising biomedical niche for the territory's scientists and entrepreneurs. Hong Kong's Innovation and Technology Commission, the main applied research funding body in the territory, has already spent more than HK\$100 million over the past few years on research into traditional medicine, involving many of Hong Kong's leading biomedical researchers.

"I'm now convinced the West could learn a lot from Chinese medicine," says Nikolaus Sucher of the Hong Kong University of Science and Technology. Sucher's laboratory has screened 22 of the some 600 materials used in Chinese traditional medicine that were selected because they may protect the brain from damage during a stroke.

Sucher's team identified compounds that block the activity of certain cell receptors — molecules in the cell membrane that receive signals and trigger responses in the cell. Excessive activity of these receptors during a stroke have been shown to kill cells. Sucher's results offer a possible explanation for the pharmacological activity of the compounds.

The new institute will start its operations by coordinating research activities and distributing grants, but will eventually erect its own building in Hong Kong's science park.

The news has attracted a lot of attention in the rest of China, says Yitao Wang, a former vice-president of the China Academy of Traditional Chinese Medicine. Wang says he knows of no such concentrated investment



Medicine men: should Hong Kong's biotechnology industry be buying into traditional remedies?

in traditional Chinese medicine anywhere else. "Given its historical and business connections to the West and its close cultural connections to China, Hong Kong could very well become the centre for internationalization of Chinese medicine," Wang says.

But some foreign and Hong Kong native researchers are less impressed. "Hong Kong is late into the race," says one visiting researcher. "Even in the United States and Europe, big pharmaceutical companies are doing research on Chinese medicine." The researcher also expresses concern over what he terms the authorities' "obsession" with traditional medicine.

Sucher and other supporters of the institute are confident that easy access to the materials and clinical histories — often available only in Chinese — will enable Hong Kong to compete with the "brute force" screening of materials carried out elsewhere. "We can be more selective," says Michael Yang of City University of Hong Kong, who is working on microarrays for screening traditional compounds.

Meanwhile, researchers in Hong Kong are jockeying for institute grants, details of which will be determined at the institute's next board meeting in August.

♦ <http://www.info.gov.hk/itc>

US blood ban underlines CJD fears

Meredith Wadman, Washington

The US government is tightening its restrictions on blood donors in an effort to prevent the spread of the human form of mad cow disease. Variant Creutzfeldt-Jakob disease (vCJD) has killed about 100 people, mostly in Britain.

Advisers to the Food and Drug Administration (FDA) met on 28 June and recommended expanding an existing blood-donor ban to include people who have spent five years or more in Europe since 1980. They also recommended banning donors who lived in Britain for three months or more between 1980 and 1996. Current restrictions, put in place in 1999, ban donors who spent six months or more in Britain during that period.

The recommendations of the Transmissible Spongiform Encephalopathies Advisory Committee are only advisory, but the agency is expected to implement them.

In January, the advisory panel

recommended that the agency ban blood donors who had spent 10 years in what it considered to be the high-risk countries of France, Portugal or Ireland since 1980 (see *Nature* 409, 441; 2001). But the FDA, unsatisfied at the failure to address risks in the rest of Europe, asked the panel to reconsider the problem.

The FDA estimates that the new restrictions will reduce the number of blood donors by about 5%. Some officials opposed the new restrictions, arguing that they would reduce an already tight blood supply.

From September, the Red Cross plans to exclude donors who have spent six months or more in Europe since 1980. The FDA says that such a move would reduce total blood donors in the United States by up to 9%.

So far, there has been no epidemiological evidence that vCJD has been transmitted by blood, but animal models suggest that this may be possible. There is no validated test that allows the screening of blood donors.

Misconduct claims derail German study of cancer vaccines

Munich Fresh allegations of scientific misconduct are rocking the German cancer-research community.

Earlier this week, the *Süddeutsche Zeitung* revealed that the University of Göttingen is investigating alleged misconduct in a project to develop a cancer 'vaccine'. The scandal comes just a year after the final report on Germany's biggest scientific fraud case, in which cancer researchers at the University of Freiburg were found to have fabricated data (see *Nature* **405**, 871–872; 2000).

The new allegations surround a project that came to prominence last year, with the publication of a paper (Kugler, A. *et al.* *Nature Med.* **6**, 332–336; 2000) reporting regression of secondary tumours. Kidney cancer patients had been treated with a vaccine composed of tumour cells fused with immune cells known as dendritic cells.

According to the *Süddeutsche Zeitung*, the investigation focuses on the conduct of Alexander Kugler, of the University of Göttingen, and Gernot Stuhler, of the University of Tübingen. It reportedly deals with allegations that a picture in another paper submitted for publication,

purportedly showing fused tumour–dendritic cells, had in fact been downloaded from the Internet. It also looks into claims of careless clinical practice and of endangering patients. The recruitment of new patients into the study has been suspended.

Bank of England opens door for life-sciences firm

London In an unprecedented move, the British government has arranged for Huntingdon Life Sciences (HLS), the drug-testing firm targeted by animal-rights protesters, to open an account with the Bank of England.

Usually, the Bank of England only handles accounts for government departments and other banks. The decision to extend its



Rights issue: banks severed ties with Huntingdon Life Sciences after animal-rights protests.

services to HLS marks the government's determination not to let protesters drive HLS out of business or out of the country.

Many banks and financial institutions have severed links with HLS following protests and intimidation from animal-rights activists (see *Nature* **411**, 7; 2001). HLS had warned that its difficulties in finding banking facilities were forcing it to consider relocating abroad.

Japan chases blue roses and better rice

Tokyo Research on genetically modified (GM) plants is quietly progressing in Japan, thanks to investment at the regional government level. Although most Japanese biotechnology companies have backed away from GM crops for fear of a public backlash, a survey by the newspaper *Asahi Shimbun* has revealed that 34 of Japan's 47 regional administrations are investing in the technology.

Projects include efforts to create improved strains of rice, blue roses and insect-resistant grass — all with the goal of boosting the competitiveness of Japanese farmers and horticulturists. The regional governments' investment is thought to total US\$3 million per year. One strain of herbicide-resistant rice, the closest project to reaching fruition,

PA PHOTOS

is being grown in fenced experimental plots and has been approved for growth in larger, open paddies in the Aichi region.

Senate intervention halts Smithsonian closure plan

Washington A key US Senate committee has waded into the debate over the future of the Smithsonian Institution, voting to reprimand a research centre that was slated for closure.

In April, staff reacted with alarm to a plan to streamline the Smithsonian's research. The plan proposed closing two of its research centres: the Conservation and Research Center in Fort Royal, Virginia, and the Center for Materials Research and Education in Suitland, Maryland (see *Nature* 410, 727; 2001).

The plan's architect, Smithsonian secretary Lawrence Small, has since backed down over closing the conservation centre. But the closure of the materials research centre, which works on techniques to preserve museum specimens and employs 29 staff, was approved in May by the Smithsonian's governing board of regents.

Both of Maryland's Democratic senators, Barbara Mikulski and Paul Sarbanes, opposed the closure. And the budget bill approved by the Senate Appropriations Committee last week maintains funding to the centre. The bill must be approved by the

full Congress and be signed by President George W. Bush before it becomes law.

Defence spin-off joins queues for contracts

London Britain's largest — and most curiously named — research company was born earlier this week, spun out of the government's Defence Evaluation and Research Agency (DERA). The new company, QinetiQ (pronounced 'kinetic'), was launched on 1 July, taking about three-quarters of DERA's 12,000 staff.

DERA researchers who work on sensitive military projects, including countermeasures against chemical and biological weapons, remain with the Ministry of Defence, in what is now called the Defence Science and Technology Laboratory. QinetiQ will compete for commercial research contracts in areas such as applied electronics and advanced materials. Its shares will initially be owned by the government, but some will later be sold to investors.

► <http://www.qinetiq.com>

Altered orbit gives voice to silent spacecraft

Munich The Huygens probe to Saturn's moon Titan, thought to have been silenced by a

design fault, will send data as planned. Huygens, built by the European Space Agency (ESA), will be jettisoned into Titan's atmosphere from the NASA-led international Cassini spacecraft.

Cassini-Huygens has been travelling to Saturn since 1997. But last year engineers discovered that the specifications of Cassini's antenna that will receive the Huygens data failed to take account of the Doppler effect. This change in frequency occurs when a source and receiver of radio waves move relative to one another. It means that nearly all of Huygens' data would have been out of the receiver's frequency range.

An ESA/NASA task force has now come up with a rescue plan. Cassini will fly much farther from Titan than planned — 65,000 km instead of 1,200 km — and on a different trajectory. This will minimize the Doppler shift, bringing the radio frequencies into the receiver's range.

Huygens' encounter with Titan will now occur on 14 January 2005, seven weeks later than planned.



Cassini-Huygens: back on course.

ESA

'Which side are you on?'

Chinese-Americans form a cornerstone of the US scientific workforce. Yet recent developments have led some to question whether they are fully accepted by their colleagues. Josette Chen reports.

Since the late 1960s, when the excesses of the Cultural Revolution caused many Chinese researchers to flee to the United States, Chinese-Americans have moved to a central position within US science. As other minorities struggled to increase their representation in US labs, Chinese-Americans were seen as an unqualified success story, with few worries about integration and discrimination.

But the arrest in December 1999 of Wen Ho Lee, a fluid-dynamicist working on nuclear-weapons research at the Los Alamos National Laboratory in New Mexico, changed the landscape. Pursuing allegations that China had acquired designs for advanced US nuclear warheads, the Federal Bureau of Investigation charged Lee with 59 counts of mishandling classified information, subjected him to repeated aggressive interrogations and held him in solitary confinement for nine months.

By the time Lee was released — having admitted one minor infringement but otherwise cleared of wrongdoing — many Chinese-American scientists were rethinking their assumptions about successful integration into the US research community. Lee, born in Taiwan but a naturalized American, was widely seen as the victim of barely concealed racism. Some Chinese-American scientists began to suspect that they, too, were perceived as 'foreigners' with divided loyalties. "History shows that Chinese-Americans get blamed for what China does," claims Dorothy Ng, who works in the Structural and Applied Mechanics Group at the Lawrence Livermore National Laboratory in California.

If such attitudes were indeed lurking below the surface, some Chinese-Americans wondered if they might be manifest in more subtle ways. Missed promotions, perhaps? Or salary discrepancies? The result was an unprecedented upsurge of activism, as Chinese-Americans sought to assert their rights. At Los Alamos and other US weapons labs, this activity was especially intense, with researchers challenging their employers' record for equal opportunities in promotions and salaries, and railing against the inclusion of 'racial profiling' within the labs' security procedures. Two academic groups, Asian Pacific Americans in Higher Education and the Association for Asian American Studies, even urged Chinese-Americans to



Touchstone issue: the arrest of Wen Ho Lee (right) raised concerns about discrimination against Chinese-American researchers.

mount an employment boycott of the labs.

A year on, *Nature's* enquiries reveal that opinion among the Chinese-American research community is divided over the progress that has been made since the issues were brought out into the open — and even over the extent to which discrimination poses a problem. But there is a general wariness over how recent developments that have strengthened the perception of China as the United States' main adversary are likely to affect attitudes towards Chinese-Americans.

Rising tension

Since President George W. Bush came to office in February, his administration has repeatedly locked horns with China, clashing over the sale of US arms to Taiwan and the dramatic incident that saw a US electronic surveillance aircraft forced to land in China after colliding with a Chinese jet fighter. In April, at the height of these tensions, an opinion poll commissioned by the Committee of 100, which lobbies for Chinese-American rights, suggested that many Americans view fellow citizens of Chinese descent with distrust.

That poll asked respondents whether they agreed or disagreed with a series of



stereotypes — and so might have been likely to provoke racist sentiments. But Ng, for one, believes some of the negative attitudes that it highlighted have spilled over into the scientific community. After the spyplane incident, she says: "I was asked at work which side I was on." Ng complains that she is expected to take a 'Chinese' view on such issues, even though she has lived in the United States for 23 years. "They have not accepted us as countrymen," she asserts.

Yet for every Chinese-American scientist who is concerned about geopolitical tensions fuelling discrimination, it seems that there is another who does not perceive a problem. "There have been ups and downs at the government level, but this has not affected my personal research area," says Zhexi Luo, a palaeontologist at the Carnegie Museum of Natural History in Pittsburgh. "Scientists are very open-minded," agrees Yixian Zheng, an embryologist at the Carnegie Institution of Washington in Baltimore, Maryland.

This divergence of views may, in part, reflect the distance of individual researchers from the centre of last year's storm — the national weapons laboratories. And when

judging the progress made since the Lee case, these labs are a good place to start.

The weapons labs are the responsibility of the Department of Energy (DOE). In summer 1999, as the storm surrounding alleged Chinese espionage at the labs grew, Bill Richardson, then energy secretary, set up a task force to examine accusations of racial discrimination in the labs' security-profiling procedures. Its recommendations, delivered in January 2000, included establishing a post of national ombudsman for the DOE to deal with such complaints. In April this year, Spencer Abraham, Richardson's successor in the Bush administration, issued a memo that pledged to eliminate racial profiling.

Slow progress

But some scientists complain that things are moving more slowly at the grass-roots level. Shao-Ping Chen, a physicist at Los Alamos, was prominent in last year's campaign to challenge the lab's administration over equal opportunities in salaries and promotion. He also sits on Los Alamos' Asian-Pacific Islander Career Enhancement Task Force, set up by lab director John Browne. "Management is moving in a positive direction, but not as fast as I'd like," Chen says.

Although Chen was promoted to a managerial position in April, he believes problems with discrimination have not been completely resolved — and he is convinced that salary discrepancies remain. In January, Los Alamos asked a consulting firm to investigate the issue. Its report, completed in May, concluded there were no significant salary differences between scientists from an Asian or Pacific Island ethnic background and other Americans. But Chen feels that the report did not provide enough data on qualifications and experience to allow meaningful comparisons.

The additional security measures imposed following the investigations into spying have also created problems. All foreign trips and contacts are now being closely monitored, and security clearance to work on classified projects has become more difficult to obtain. Chen is himself still waiting for clearance after applying some 18 months ago. Chen is not sure if his status as a Chinese-American has hindered the process — but he says that anyone with contacts in 'sensitive' countries will be looked at more closely.

Lawrence Livermore, meanwhile, faces a lawsuit filed by nine Asian-American scientists and engineers alleging bias in hiring, salaries and promotions. The plaintiffs claim there is a salary gap averaging \$922 per month between them and their white counterparts.

Calls for an employment boycott have also taken their toll, with appointments of Asians and Asian-Americans to postdoctoral positions at the weapons labs falling from 14% of the total in 1998 to 7% in 2000. Ling-chi Wang, an ethnic studies professor at the University of California, Berkeley, who led calls

for a boycott, believes it has prompted some progress. But he fears this has been "blown away" by a recent incident involving Representative David Wu (Democrat, Oregon).

In May, Wu was invited to speak at the DOE's headquarters to celebrate Asian Pacific American Heritage Month. Turning up at a side entrance, he was briefly denied access — despite his congressional identification — because he could not prove his US citizenship. In a formal complaint, Wu told energy secretary Abraham that he suspected the incident was "an indicator of a much larger problem at DOE", and has asked for a review of employment practices and operating procedures to prevent future discrimination.

David Ho, director of the Aaron Diamond AIDS Research Center in New York and a member of the President's Advisory Commission on Asian Americans and Pacific Islanders, believes that such problems are less acute in the wider scientific community. But again, he is concerned about the impact recent tensions between the United States and China will have on attitudes towards Chinese-American scientists.

There are also concerns about the flow of young scientists from China into US research labs. This is widely felt to benefit both countries, and some researchers fear that recent events may deter bright young Chinese from visiting the United States. "If we lose them, it won't be good for US science," warns Cheuk-Yin Wong, a nuclear physicist at the Oak Ridge National Laboratory in Tennessee and former chairman of the Overseas Chinese Physics Association.

Away from the weapons labs, at least, there seems to be no cause for immediate alarm. Nai Phuan Ong, a physicist at Princeton University in New Jersey, says that student admissions in his department have not been affected. Given political tensions, Chinese scientists "may feel bad or worry" about

There is a general wariness over how recent events will affect attitudes towards Chinese-Americans.



Ground zero: the Los Alamos National Laboratory has been accused of 'racial profiling'.

coming to the United States, he says. But the draw of working in the world's leading scientific nation remains strong — and many young Chinese researchers are in contact with friends working in the United States who can relate positive experiences.

As Chinese-Americans worry about their status within US society, politicians are reacting to their concerns. Last month, Senator Dianne Feinstein (Democrat, California) drew attention to "the growing web of suspicion" surrounding Asian-Americans and asked President Bush to "redouble our efforts to eliminate racial stereotypes".

Such stereotypes, argue Chinese-American scientists, are as unwelcome in the lab as anywhere else. As the Carnegie Museum's Luo puts it: "Truly good science should have no boundaries, no ethnic attitudes."

■ Josette Chen has just completed an internship at Nature.



Sour times: the forced landing of a US surveillance aircraft in China sparked an international incident.



Meet the Herod bug

A parasitic bacterium that uses an array of dastardly tricks to favour female hosts over males is holding evolutionary biologists in thrall. Jonathan Knight enters the strange world of *Wolbachia*.

According to the Bible, King Herod had the blood of thousands of male children on his hands. Yet his record for gender-biased genocide pales alongside that of *Wolbachia*, bacteria that infect a wide variety of invertebrates including insects, spiders, crustaceans and nematode worms.

The bacteria, which live inside the cells of its hosts' reproductive systems and other tissues, can only spread from one generation to the next by invading a female's eggs. Males, to *Wolbachia*, are useless. Accordingly, these bacteria remove males from the populations they infect with a chilling efficiency, using strategies that range from enforced sex changes to cold-blooded murder.

Wolbachia were long regarded as a biological quirk. But now it seems that they are much more prevalent than was originally thought, and evolutionary biologists are starting to view these bizarre parasites as a unique natural laboratory. Thanks to their male-hating habits, *Wolbachia* might help investigations of such key questions as how species are created and go extinct, and why different organisms determine the difference between the sexes in myriad ways. "It's clear that they are a major

force in the biology of a lot of invertebrates," says Scott O'Neill, who studies the bacteria at Yale University in New Haven, Connecticut.

Wolbachia disrupt their hosts' biology in many ways. In mosquitoes, for instance, infected males can only reproduce successfully with females infected with the same strain of *Wolbachia*¹. Although the bacteria themselves cannot hitch a ride in sperm, they alter the cells by an unknown mechanism so that the sperm can only fertilize eggs that contain the same *Wolbachia* strain. This 'cytoplasmic incompatibility' reduces reproduction among uninfected females, and so drives the infection through the population with each generation. *Wolbachia*-induced cytoplasmic incompatibility has also turned up in beetles², wasps³, moths⁴ and fruitflies⁵.

In other species, *Wolbachia* makes males dispensable. Some populations of parasitic wasp in the genus *Trichogramma* are entirely female, reproducing parthenogenetically — without fertilization. Parthenogenesis arises from time to time in animals ranging from insects to fish and lizards. But in 1990, Richard Stouthamer of the University of California, Riverside, found that antibiotics

could 'cure' parthenogenesis in *Trichogramma*⁶. Later, he revealed that the culprit was a *Wolbachia* infection⁷.

Wolbachia also convert male woodlice into females, apparently by suppressing a gland that produces a masculinizing hormone⁸. In species including ladybirds, flour beetles and several species of fruitfly, *Wolbachia* simply kill males during early development^{9–11}. By preventing these 'dead-end' hosts from hatching, they ensure better supplies of food for their infected daughters. "It's a bacterium that can do it all," says Stouthamer.

Out of the shadows

Wolbachia were relatively obscure until John Werren of the University of Rochester in New York looked for them in the wild — and found them everywhere. Using the polymerase chain reaction (PCR) to copy fragments of bacterial DNA from insect tissue samples, Werren discovered evidence of *Wolbachia* infection in about 17% of the species he surveyed in Panama and 20% of those in the United States and Canada¹².

Wolbachia might in fact be even more widespread than that. Last year, Marjorie



Hitching a ride: *Wolbachia*, seen here as bright spots in an insect egg, pass from one generation of their host to the next.



Victimized: *Wolbachia* distort the biology of a wide range of insects including ants, termites, dragonflies and cockroaches.



Jay Jeyaprakash (left) and Marjorie Hoy's studies of insects, including this ladybird, led them to conclude that *Wolbachia*'s prevalence is underestimated.



Hoy of the University of Florida in Gainesville was working with a species of spider mite that had a skewed sex ratio highly suggestive of a *Wolbachia* infection. She could not reproducibly detect the bacterium with standard PCR. Only 'long' PCR, which includes a second enzyme to correct errors in copying the DNA, gave consistent results.

This led Hoy to suspect that Werren's study might have underreported *Wolbachia*'s incidence. She teamed up with her Florida colleague Jay Jeyaprakash and collected specimens from 61 insect species and two species of spider from the field or local laboratory stocks. "We got some cockroaches and house flies and termites and dragonflies and ants — whatever was available," says Hoy. With long PCR, 76% of the species tested positive for *Wolbachia*¹³.

Some researchers want to see further evidence before accepting that *Wolbachia* are so remarkably widespread. "I'd like to see that work replicated," says Francis Jiggins of the University of Cambridge. But Jiggins has his own evidence that Werren's estimate is low.

Werren only sampled one or two individuals from each species he tested. If *Wolbachia* only infects a small proportion of a population, he may have missed many species that do carry the microbe. Jiggins reported last month that, at least for butterflies, this is the case¹⁴. He surveyed a large number of individuals within a small number of African butterfly species, and found that typically only 15% of the individuals are infected. "We can certainly say that Werren's is a conservative estimate," Jiggins concludes.

If *Wolbachia* are as prevalent as these studies suggest, then the bacteria are likely to exert a powerful influence over the evolution of their hosts. One provocative proposal is that they might drive speciation — the emergence of new species. Speciation begins with the reproductive isolation of two populations of the same species. A mountain range, for instance, could separate populations for millennia until the accumulation of muta-

tions in each prevents them from producing viable offspring if they ever do meet.

Cytoplasmic incompatibility could provide a similar reproductive barrier. If two groups within a population of insects are infected with two different types of *Wolbachia*, males from either group may be unable to fertilize females from the other. Werren believes such 'bi-directional' cytoplasmic incompatibility might allow speciation to occur. The best evidence so far comes from two species of wasp, *Nasonia longicornis* and *N. giraulti*. They look different — *N. giraulti* has much smaller wings — and do not interbreed. But they are closely related: DNA analyses suggest that their most recent common ancestor lived around 250,000 years ago¹⁵.

In February, Werren and his graduate student Seth Bordenstein reported that the two species were infected with different strains of *Wolbachia* that were causing bi-directional cytoplasmic incompatibility¹⁶. Once cured with antibiotics, the wasps could produce fertile offspring. Infection with different strains of *Wolbachia* appeared to be one of the first reproductive barriers to have arisen between these recently divergent species.

Divide and conquer

But some experts question whether *Wolbachia*-induced cytoplasmic incompatibility is, by itself, enough to cause species to diverge. Scott Turelli, an evolutionary biologist at the University of California, Davis, points out that Werren's two wasp species are geographically separated: *N. giraulti* lives in eastern North America, whereas *N. longicornis* inhabits the western part of the continent. And Stouthamer notes that cytoplasmic incompatibility can break down, for instance where insects mate repeatedly¹⁷. He believes this may allow enough gene flow to prevent co-existing populations from becoming reproductively isolated. "My feeling is that cytoplasmic incompatibility can help once speciation has started, but it can't initiate speciation," he says.

John Jaenike, an evolutionary biologist at the University of Rochester, may have stumbled upon just such a case among fungus-eating flies. One species, *Drosophila recens*, collected from the eastern United States, is heavily infected with *Wolbachia*. Its males cannot fertilize females from another species, *D. subquinaria*, that lives in the Pacific northwest and is *Wolbachia*-free. The reverse cross — *D. subquinaria* males and *D. recens* females — does produce offspring. But in unpublished observations, Jaenike has found that *D. recens* females are reluctant to mate with their western relatives. He has yet to study the flies' behaviour in the zone in the central United States where their ranges

Wolbachia are likely to exert a powerful influence over the evolution of their hosts.

overlap, but it is possible that female mating preference prevents hybridization in one direction whereas *Wolbachia*-induced cytoplasmic incompatibility stops it in the other.

Determined response

Speciation is not the only aspect of evolution in which *Wolbachia* may have a hand. Greg Hurst of University College London believes the bacteria may have helped to spawn the tremendous diversity in the way in which different invertebrate species determine whether to be male or female.

One method of sex determination, found in *Drosophila*, depends on the ratio of X chromosomes to autosomes (all the other chromosomes except the Y chromosome). Flies with two X chromosomes and two copies of each autosome will be female, whereas a 1:2 ratio produces a male. In *Trichogramma* wasps, fertilized eggs become female, and unfertilized eggs become male — unless a *Wolbachia* infection forces them to develop as females. There are many other mechanisms in different species.

To Hurst, this is an intriguing puzzle. In other respects, sex is relatively invariant — males produce sperm and females make eggs. So why should sex determination be different? "One hypothesis is that it's driven by parasites that affect sex," says Hurst. In other words, the struggle between the hosts' attempts to produce an even sex ratio, and *Wolbachia*'s desire for more females than males, has led hosts to evolve a variety of means to influence sex.

In March, for instance, Stouthamer's group reported the discovery of a 'parasitic'



Raising awareness: until John Werren looked for *Wolbachia* (inset inside a host cell) in the wild, they were seen as an obscure oddity.



Natural wonder: *Wolbachia*'s diverse effects on its hosts have impressed Richard Stouthamer.

chromosome in *Trichogramma* which causes females to produce only males¹⁸. This chromosome appears to restore some balance between the sexes to *Wolbachia*-infected wild populations of the wasp. 'Masculinizing' genes have also been found in infected woodlice¹⁹.

Decline and fall

For most species, a distorted sex ratio makes reproduction less efficient. But in theory, if the distortion becomes severe enough that females struggle to find mates, it could drive populations — and even species — to extinction. Jiggins and Hurst are investigating this possibility. Two years ago, they reported cases of male-killing by *Wolbachia* in ladybirds from Russia and butterflies from Uganda²⁰. Both populations were more than 90% female, and nearly all individuals were infected.

Wolbachia has so skewed the gender ratio of the Ugandan butterfly, *Acraea encedon*, that a rare form of courtship has arisen. Usually, males compete for the attentions of choosy females. But in *A. encedon*, females wait in groups for a male to flutter by and then vie for a chance to mate with him. Many females never mate. "The question is whether they could ultimately go to extinction," says Hurst. Jiggins and Hurst are now separately studying similarly afflicted insect populations for evidence that some truly are on the brink.

A knock-on effect of male-culling might be to restrict where an insect species can live. For reasons that are not understood, *Wolbachia* do not transfer to the next generation as efficiently at higher temperatures²¹. At low temperatures, it may transmit so effectively that its male-killing habit can make populations extinct. So a population infected with *Wolbachia* might only survive in warmer climates. Jaenike is now testing this idea in *Drosophila innubila*, which lives in Arizona. He wants to know if male-killing *Wolbachia* are restricting the flies to life at warmer, lower elevations.

Intriguingly, just as *Wolbachia* are integral to the biology and evolution of their

hosts, it seems that a virus known as the WO phage may be integral to the biology of *Wolbachia*. Shinji Masui and his colleagues at the University of Tokyo last year reported finding phage DNA sequences embedded in the chromosomes of several *Wolbachia* strains²².

Phages can hitch a ride from their bacterial host for generations by inserting their DNA into the host's genome. Later the phage hops back out, gets packaged in its protein coat and bursts out of the cell to infect more bacteria. Sometimes, it takes a bit of bacterial DNA with it, and Masui suspects that this may allow the WO phage to shuttle genes from one *Wolbachia* strain to another. If so, it could explain why closely related strains of *Wolbachia* can cause radically different effects on their hosts. Genes involved in male killing, feminization and parthenogenesis might have travelled from one species to another many times, mixing up the *Wolbachia* evolutionary tree.

Preliminary evidence that genes are jumping between *Wolbachia* strains is now emerging from several labs that are sequencing the *Wolbachia* genome. O'Neill's Yale group has nearly finished one *Wolbachia* strain from a nematode worm, *Brugia malayi*, that lives in the human lymph system and blood, and another from *Drosophila melanogaster*. Although the sequences are still being analysed, there appear to be many mobile genetic elements such as phage DNA and transposons, 'jumping genes' that can hop in and out of the genome, inserting randomly.

Two become one

Wolbachia has a small genome — less than 1.5 million base pairs, or about a third of the size of the genome of the common gut bacterium *Escherichia coli*. Intracellular parasites often seem to develop slimmed-down genomes, as they lose genes that are only required by free-living bacteria²³. And some biologists wonder if *Wolbachia* are on the evolutionary road towards becoming one with their hosts — like the mitochondria that generate energy inside our cells, which are thought to have once been parasites.

O'Neill, for one, is not convinced. Given the deleterious effects of *Wolbachia* on its insect hosts, he definitely views the bacteria

as parasites: "The microbe has the upper hand." But in other invertebrates, who has the upper hand is less clear. The nematode worms that cause elephantiasis and river blindness in humans seem to need *Wolbachia* to survive — antibiotics kill the worms, apparently by killing their *Wolbachia*²⁴.

As interest in the bacteria explodes, strains of *Wolbachia* that are in the process of being incorporated by their hosts may be among the evolutionary treasures waiting to be discovered. The availability of the *Wolbachia* genome is also expected to aid research into the mechanisms by which the bacteria manipulate their hosts — which at present are not well understood. Fans of bizarre biology and aficionados of evolution should watch this space.

Jonathan Knight writes for Nature from San Francisco.

1. Yen, J. H. & Barr, A. R. *Nature* **232**, 657–658 (1971).
2. Wade, M. J. & Stevens, L. *Science* **227**, 527–528 (1985).
3. Richardson, P. M., Holmes, W. P. & Saul, G. B. *J. Invertebr. Pathol.* **50**, 176–183 (1987).
4. Brower, J. H. *Ann. Entomol. Soc. Am.* **69**, 1011–1015 (1975).
5. Hoffman, A. A., Turelli, M. & Simmons, G. M. *Evolution* **40**, 692–701 (1986).
6. Stouthamer, R., Luck, R. F. & Hamilton, W. D. *Proc. Natl. Acad. Sci. USA* **87**, 2424–2427 (1990).
7. Stouthamer, R., Breeuwer, J. A. J., Luck, R. F. & Werren, J. H. *Nature* **361**, 66–68 (1993).
8. Rigaud, T., Souty-Grosset, C., Raimond, R., Mocquard, J. P. & Juchault, P. *Endocytobiosis Cell Res.* **7**, 259–273 (1991).
9. Majerus, T. M. O., von der Schulenburg, J. H. G., Majerus, M. E. N. & Hurst, G. D. *D. Insect Mol. Biol.* **8**, 551–555 (1999).
10. Fialho, R. F. & Stevens, L. *Proc. R. Soc. Lond. B* **267**, 1469–1473 (2000).
11. Hurst, G. D. D., Johnson, A. P., von der Schulenburg, J. H. G. & Fuyama, Y. *Genetics* **156**, 699–709 (2000).
12. Werren, J. H., Guo, L. & Windsor, D. M. *Proc. R. Soc. Lond. B* **262**, 147–204 (1995).
13. Jeyaprakash, A. & Hoy, M. A. *Insect Mol. Biol.* **9**, 393–405 (2000).
14. Jiggins, F. M., Bentley, J. K., Majerus, M. E. N. & Hurst, G. D. D. *Proc. R. Soc. Lond. B* **268**, 1123–1126 (2001).
15. Campbell, B. C., Steffen-Campbell, J. D. & Werren, J. H. *Insect Mol. Biol.* **2**, 225–237 (1993).
16. Bordenstein, S. R., O'Hara, F. P. & Werren, J. H. *Nature* **409**, 707–710 (2001).
17. Hoffmann, A. A., Hercus, M. & Dagher, H. *Genetics* **148**, 221–231 (1998).
18. Stouthamer, R., van Tilborg, M., de Jong, J. H., Nunney, L. & Luck, R. F. *Proc. R. Soc. Lond. B* **268**, 617–622 (2001).
19. Rigaud, T. & Juchault, P. *Genetics* **133**, 247–252 (1993).
20. Hurst, G. D. D. *et al. Proc. R. Soc. Lond. B* **266**, 735–740 (1999).
21. van Opijnen, T. & Breeuwer, J. A. *Exp. Appl. Acarol.* **23**, 871–881 (1999).
22. Masui, S., Kamoda, S., Sasaki, T. & Ishikawa, H. *J. Mol. Evol.* **51**, 491–497 (2000).
23. Cole, S. T. *et al. Nature* **409**, 1007–1011 (2001).
24. Taylor, M. J., Bandi, C., Hoerauf, A. M. & Lazdins, J. *Parasitol. Today* **16**, 179–180 (2000).



Sexually harassed: *Wolbachia* kills male ladybirds, makes *Trichogramma* wasps (centre) reproduce asexually and has made populations of *Acraea encedon* butterflies 90% female.

Identifying dangers in an uncertain climate

We need to research all the potential outcomes, not try to guess which is likeliest to occur.

Sir — Stephen Schneider, in his Commentary “What is ‘dangerous’ climate change?” (*Nature* **411**, 17–19; 2001), provides a succinct description of the IPCC process and an excellent summary of the Special Report on Emissions Scenarios (SRES, of which we were among the lead authors), its findings and implications for climate change. He also gives a perceptive account of the critical issues that remain unresolved after completion of the comprehensive Third Assessment Report of the IPCC (see Schneider’s Commentary for publication references).

Although we agree with most of what Schneider says, we disagree with him about the appropriateness and feasibility of assigning subjective probabilities of occurrence to alternative, unknown futures described by the SRES scenarios. In an interdisciplinary scientific assessment, the concept of probabilities as used in natural sciences should not be imposed on the social sciences. Probability in the natural sciences is a statistical approach relying on repeated experiments and frequencies of measured outcomes, in which the system to be analysed can be viewed as a ‘black box’. Scenarios describing possible future developments in society, economy, technology, policy and so on, are radically different.

First, there are no independent observations and no repeated experiments: the future is unknown, and each future is ‘path-dependent’: that is, it results from a large series of conditionalities (‘what if... then’ assumptions) that need to be followed through in constructing internally consistent scenarios. Socio-economic variables and their alternative future development paths cannot be combined at will and are not freely interchangeable because of their interdependencies. One should not, for example, create a scenario combining low fertility with high infant mortality, or zero economic growth with rapid technological change and productivity growth — since these do not tend to go together in real life any more than they do in demographic or economic theory.

Second, the ‘what if... then’ approach requires the explicit representation of a system to describe how the variables interact. That is, any randomly assumed distribution of ‘what if’ conditions (based on expert opinion about the future of world population, economic growth, technological development, diets and so on) would be insufficient without also

assigning distributions to the conditional probabilities by which these assumed future states of driving forces interact to influence an unknown outcome (greenhouse-gas emissions). This means that the distributions of ‘what if’ as well as ‘then’ remain unknown.

We agree with Schneider about the dangers of arbitrarily picking a (too) limited set of scenarios. We also agree that it is important to be consistent in choosing which climate sensitivities to apply to which scenario. Resorting to a premature ‘expert consensus’ on the likelihood of certain emission futures and associated climate change in 100 years seems to ignore two essential facts: climate change has moved beyond the domain of ‘experts’ to that of a multitude of societal stakeholders; and good scientific arguments preclude determining

‘probabilities’ or the likelihood that future events will occur. The levels of future greenhouse-gas emissions and the ensuing climate change remain uncertain; we need adaptive response strategies that explicitly recognize these uncertainties.

So although we agree with Schneider in many respects, ‘dangerous’ levels of climate change will need to be identified by research into the adverse impacts on natural and human systems, independent of the question of how likely they are to occur, and covering the full range of scientific uncertainty.

There is a danger that Schneider’s position might lead to a dismissal of uncertainty in favour of spuriously constructed ‘expert’ opinion.

Arnulf Gröbler, Nebojsa Nakicenovic
International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Attention to bioweapons obscures the real threats

Sir — Your News Feature “The bugs of war” (*Nature* **411**, 232–235; 2001) brings lucidity to the debate about the threat of biological weapons. Until now, most assessments of this issue have been woefully inadequate in that they have dealt exclusively with political evaluations of possible technical vulnerabilities. Your News Feature gives an up-to-date depiction of the scientific reality behind the lopsided policy deliberations. We would like to add some points.

First is the issue of political motivation. If genetically engineered bioweapons were developed, would they be deployed by nation-states, terrorists or criminals? The problem is that as things stand there are more than enough non-engineered biological agents available to satisfy all these categories. Second, technical expertise is far too often overlooked. Even if a state or non-state actor has the wherewithal to manufacture an engineered bioweapon — which entails extensive technical expertise and a huge budget — there remains the severe problem of adequate dispersal necessary for mass casualties. Bioweapons cannot be spoon-fed to the intended targets. Delivery is an exceptionally difficult science, as has been found in state programmes (for example in the United States, the Soviet Union and Iraq) to create bioweapons. Finally, in any

assessment it is important not to fall into the anthropocentric trap. The dangers of bioweapons directed at livestock and crops are as great as, if not greater than, those designed against people.

The conclusion of your News Feature, that whatever advances are made on the offensive side will have beneficial spin-offs on the defensive side, is important. Despite the overwhelmingly positive influence on society of the biomedical revolution, especially in genomics and proteomics, there are still problems, their potential misuse for weapons being just one. Knowledge does not automatically lead to capability. Your Opinion article in the same issue, “A call to arms” (*Nature* **411**, 223; 2001) reminds us that, for terrorists, the gun and the bomb will continue to be the weapons of choice simply because these are more than adequate tools for political violence. As a result, we should be wary of being seduced by talk of a ‘threat’ of super-bioweapons. A balanced response by governments relies on accurate and proportionate risk assessments of the bioweapon threat. Without these, time and money will be wasted and, perhaps more important, the real threats will be overlooked.

Sebestyén Gorka*, Richard Sullivan†

**US Council for Emerging National Security Affairs, PO Box 378 Solebury, Pennsylvania 18963, USA*

†Scientific Department, Cancer Research Campaign, 10 Cambridge Terrace, London NW1 4JL, UK

Quest for the ancient fire-breathers

Those who pursued dinosaurs could be as fascinating as their discoveries.

Dragon Hunter: Roy Chapman Andrews and the Central Asiatic Expeditions

by Charles Gallenkamp
Viking: 2001. 432 pp. \$29.95

The Dragon Seekers: How an Extraordinary Circle of Fossilists Discovered Dinosaurs and Paved the Way for Darwin

by Christopher McGowan
Perseus: 2001. 256 pp. \$26

Eric Buffetaut

With the exception of Thomas Hawkins — an eccentric British collector of Jurassic marine reptiles — the men and women described in these two books for the general reader probably did not think of themselves as the discoverers of dragons. But the metaphor is engaging: dinosaurs and other 'prehistoric monsters' may indeed play the part of the dragons of old in the modern collective unconscious. And those who seek the dragons of the past are sometimes as fascinating as the creatures they discover, as both books make clear.

Roy Chapman Andrews — Charles Gallenkamp's 'dragon hunter' — was a real-life big-game hunter who killed hundreds of specimens of the wildlife of Asia for the American Museum of Natural History in New York, the institution to which he devoted his life. Andrews was more than a mere collector, however; he was also an expert on living whales, a New York socialite who spent a large part of his life in Asia and an adventurer who professed that "adventures are a mark of incompetence". But he became world



Big-game hunter: Andrews (left and below) was famous for his dinosaur discoveries in the Gobi Desert.

famous for collecting dinosaurs in the Gobi Desert.

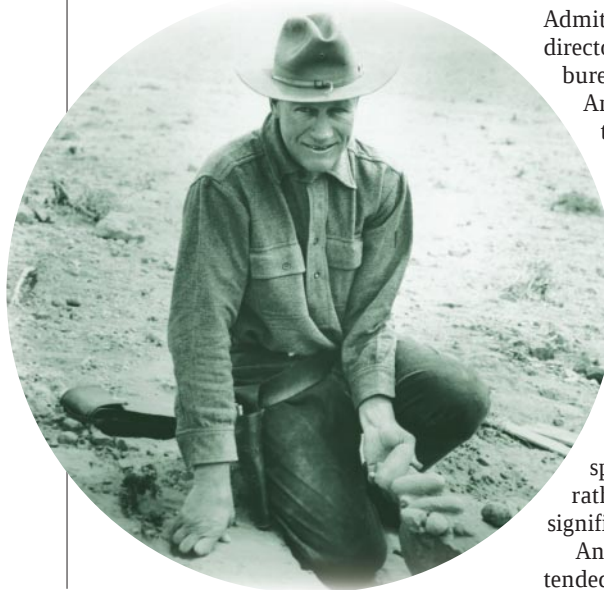
To some extent, the biography of this paradoxical character reads like an American success story, with the bright young man from Beloit, Wisconsin, rising from a menial position in the taxidermy department of the American Museum to become its most celebrated explorer, and eventually its director. Admittedly, he was not a very successful director — changing from adventurer to bureaucrat proved impossible.

Andrews' Central Asiatic expeditions of the 1920s have taken on an almost mythical dimension at the American Museum (and in popular books on dinosaurs), and they occupy a large part of Gallenkamp's book. The adventures of the Gobi expeditions have been told many times, however, even by Andrews himself, and Gallenkamp's sprightly narrative provides little that is new. And the scientific side of the story is a bit sketchy, with the accent on the spectacular aspects of the fossil finds rather than their deeper palaeontological significance.

Andrews' achievements in the Gobi have tended to overshadow his escapades before

and after, and in this respect Gallenkamp's biography is a precious source of information. Andrews' research on whales took him first to Japan, where he fell under the spell of Asia, and then to Korea and China, where, at the end of the First World War, he collected both zoological specimens for the museum and intelligence for the US government while planning his Central Asiatic expeditions.

The expeditions, with their large-scale use of motor cars, heralded a new era in scientific exploration and were also the swansong of a certain type of expedition from Western countries that paid little heed to the wishes of the local authorities. The demands of the Communist government in Mongolia and the Nationalist government in China, which so infuriated Andrews and his mentor Henry Fairfield Osborn, appear perfectly justified by present-day standards. The return of specimens to their country of origin and the training of local experts are now part and parcel of all cooperation agreements. Interestingly, although this seems to have escaped Gallenkamp's attention, the Chinese participants Andrews reluctantly took along on his final expedition were no other than C. C. Young and W. C. Pei, who were to become the founding



BETTMANN/CORBIS

BETTMANN/CORBIS



Sea dragon: skull of an ichthyosaur collected by Mary Anning in Lyme Regis, southwest England.

fathers of Chinese vertebrate palaeontology and palaeoanthropology in the 1940s and 1950s.

A century before Andrews' expeditions, a group of British "fossilists", as Christopher McGowan calls them, engaged in scientific adventures that were probably more intellectual and less physical than Andrews' trips to the Gobi — although excavating a large ichthyosaur on a Dorset beach during a winter gale may have been as daunting as digging up dinosaur eggs in the midst of a Gobi sandstorm.

Detailed biographies of many (although not all) of the protagonists in this story have been published in recent years, from William Buckland, the first professor of geology at Oxford University, to Richard Owen, Britain's equivalent of Georges Cuvier, the celebrated French palaeontologist, and the surgeon-cum-palaeontologist Gideon Mantell. McGowan's book uses a different approach: it offers the reader a reconstruction of the scientific climate in England during the first decades of the nineteenth century, and of the interactions between a group of remarkable people. They came from very different backgrounds, and were not only the discoverers of the giant reptiles of the past, but also brought about far-reaching changes in our conception of the history of life on Earth.

This is not just another dinosaur book. McGowan, himself an expert on marine reptiles, rightly recognizes that, in those days, the ichthyosaur and the plesiosaur, Hawkins' "great sea dragons", captured the attention of both scientist and layman as much as did the more poorly known dinosaurs. Beyond their spectacular discoveries, the fossilists were actors in a fundamental debate about the meaning of the fossil record and the nature of geological phenomena. Buckland's cave researches and interest in the biblical Flood receive the attention they deserve, as does Charles Lyell's victorious defence of uniformitarianism — the principle that the Earth was shaped by the same sort of forces that are still at work today.

The boundaries between professional

scientists and amateurs were not as distinct then as they are today: both Buckland and Owen held official positions, but, whereas Lyell, a lawyer by training, was a gentleman of independent means, Mantell had to practise surgery to make an often precarious living. As for Hawkins, he was apparently both a mental case and something of a scoundrel, who overpriced the sometimes unduly restored specimens he sold to the British Museum.

In many ways, the most remarkable of these fossilists was Mary Anning of Lyme Regis. Despite modest origins and a limited education, by finding remarkable fossils and selling them to wealthy collectors and institutions, she managed to win recognition in scientific circles to which women had virtually no access.

In the book's final chapter, McGowan pays a well-deserved tribute to the present-

day amateur fossil collectors who scour the cliffs of the English West Country for remains of ancient 'dragons', and thus keep alive the spirit of Mary Anning. Indeed, if a common lesson is to be drawn from these highly readable books, it is that, whatever its conceptual advances, the development of palaeontology will always depend on new fossil discoveries, whether made by professional explorers in remote parts of the Earth or by amateur fossilists on their doorsteps. Dragons are the stuff palaeontology is made of. ■

Eric Buffetaut is at the Centre National de la Recherche Scientifique, 16 cour du Liégat, 75013 Paris Cedex 05, France.

Delving into the material world

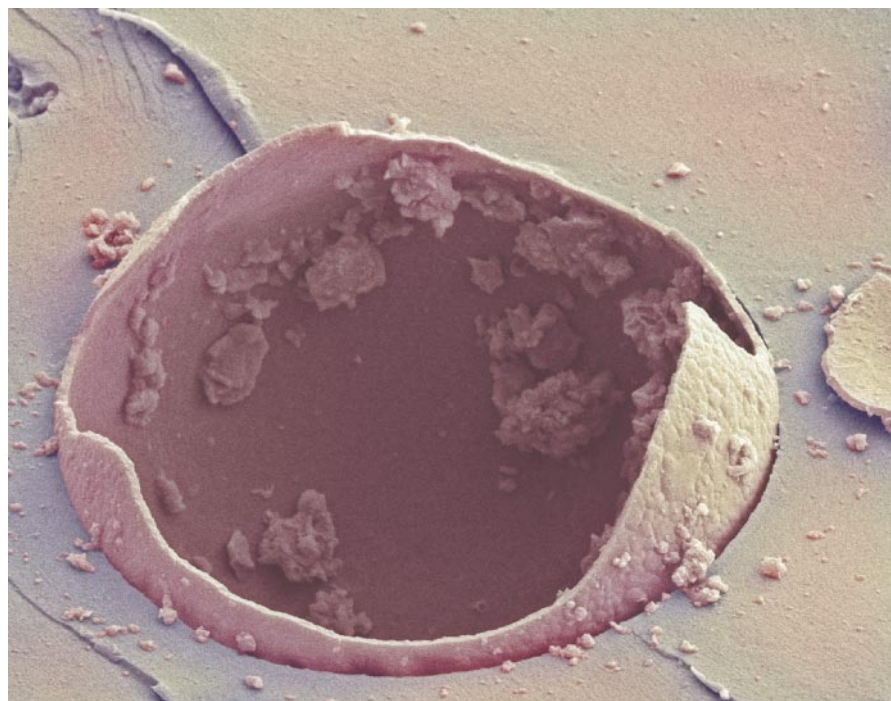
The Coming of Materials Science

by R. W. Cahn

Pergamon: 2001. 584 pp. \$64

Jacques Livage

One of the great landmarks in human development was the discovery that man could change the very nature of materials. The earliest ages of civilization are classified by the materials in use at the time — stone, copper, bronze and iron. Materials do indeed have a profound influence on the cultural, economic, demographic and geographic evolution of society. The development of aeronautics and electronics during



Resourceful resource: self-repairing plastic seen under the scanning electron microscope..



A matter of sport: section through a fibreglass-containing plastic, used to make skis and surfboards.

the twentieth century, for instance, would not have been possible without aluminium and silicon.

Ceramics were the earliest structurally modified inorganic materials to be produced by humans, and almost all the properties of modern materials, except for electrical and magnetic properties, were exploited by the early ceramicists. But, surprisingly, materials remained a craft rather than a science until the beginning of the twentieth century, and the field of 'materials science' emerged in US universities only during the early 1950s. The discipline was not introduced in Britain until well into the 1960s and even later on the European continent.

Materials science is a concept born of metallurgy, not ceramics, and Robert Cahn, a well-known metallurgist, is one of the pioneers in its development. He was the first to teach the subject in Britain and has chaired the editorial board of the *Journal of Materials Science* since its creation in 1965. *The Coming of Materials Science* is therefore not just a history of the discipline, but a personal, sometimes provocative, view of this multi-disciplinary field.

Cahn's book is primarily directed at professional materials scientists and engineers, but it could as well be read by non-specialists interested in the history of science. It follows the evolution of the science of materials from the very beginning, showing how new ideas and concepts emerged.

This is not a textbook of materials science, and should rather be regarded as a pointillist portrait of the discipline, to be viewed from a slight distance. Cahn does not attempt to give a comprehensive account of the field, but instead focuses on a few historically important developments. The book is remarkably well documented, and describes the role of important scientists,

together with discussions of the scientific concepts involved.

Northwestern University in Illinois was the first to adopt materials science as part of a department title, changing both university teaching and the organization of US academic research. Materials science grew out of departments of metallurgy, and for a long time there were departments of 'materials science and metallurgy'. Later came departments of 'materials science and engineering', emphasizing the fact that materials science in universities cannot be separated from the industrial development of materials. Some of the field's best research scientists and Nobel prizewinners worked in the industrial research labs of companies such as Bell Telephone and General Electric.

The emergence of materials science as a discipline required the understanding of a triumvirate of basic sciences: atoms and crystals, phase equilibria and microstructure. According to the author, microstructure is the component that best distinguishes 'materials science and engineering' from other disciplines. The materials scientist must work at different levels of organization, from microstructure to mesostructure, with each level being underpinned by the level before it. Each stage, from synthesis and structure to properties and processing, must be carefully controlled to optimize performance.

A quantitative theory of materials science was progressively built up as physical metallurgy developed, and the real revolution occurred around the middle of the twentieth century, leading the way to modern materials science.

This quantitative revolution owed everything to the availability of sensitive and precise techniques of measurement

and characterization, such as electron microscopy, spectrometry and diffraction methods. But materials are not limited to metals. Since the beginning of the twentieth century, materials science has developed along several separate channels — solid-state physics, metallurgy, polymer chemistry, inorganic chemistry, mineralogy, glasses and ceramics — and the discipline now includes all of these fields. The book describes these aspects in separate chapters. But, curiously, composite materials have been forgotten; the subject index gives only one page referring to composites.

Materials science is continuously evolving, and after gathering up different disciplines, it is now separating into new fields. Chemistry, for instance, plays a large part. Polymer science would not, of course, have been possible without the chemical synthesis of long-chain molecules such as polyethylene and nylon. Inorganic materials have been the main field of 'solid-state chemists' since the 1950s. New fields are still emerging; there is, for example, chimie douce ('soft chemistry'), molecular materials — which investigates novel materials whose properties originate at the molecular level — and biomimetics, which aims to apply biological designs to man-made materials and structures. We can expect even more discoveries in the twenty-first century. ■

Jacques Livage is at the Université Pierre et Marie Curie, 4 place Jussieu, 75252 Paris Cedex 05, France.

Cajal made clear in translation

Texture of the Nervous System of Man and Vertebrates: Volume II by Santiago Ramón y Cajal, translated and edited by Pedro Pasik and Tauba Pasik
Springer: 2000. 680 pp. \$150
Edward G. Jones

The Spanish neurohistologist Santiago Ramón y Cajal has not always been well served by his translators. Those translating his works into German or French during his lifetime (1852–1934) were not Spanish speakers and relied on their knowledge of Latin. They invariably communicated with Cajal in French, a language he could read but in which he could not adequately express himself.

Although those translations contain few major factual errors, terms that Cajal would not have used inevitably crept in, together with nuances of emphasis reflecting the translator's own perspective. The famous 'double bouquet cell' of the cerebral cortex, for example, is not Cajal's term but that of Leon Azoulay, who was responsible for the

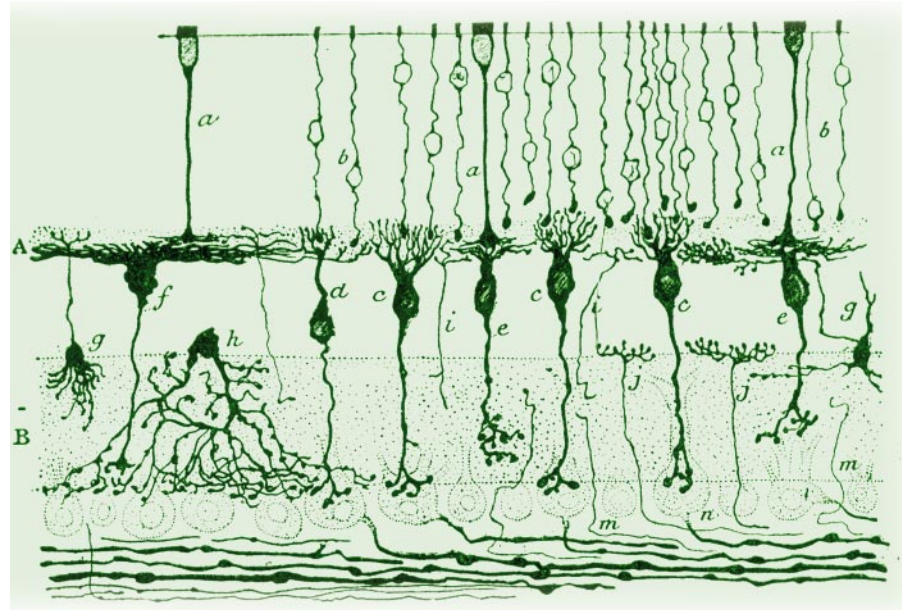
1909–11 French translation of Cajal's *Textura del Sistema Nervioso del Hombre y de los Vertebrados* (1899, 1904) as *Histologie du Système Nerveux de l'Homme et des Vertébrés*. And Cajal's florid style, with its characteristically long and involved sentences in which sometimes even the author loses his way, completely disappeared.

Early English translations fared little better. The 1928 version of Cajal's *Estudios sobre la Degeneración y Regeneración del Sistema Nervioso* (*Degeneration and Regeneration of the Nervous System*) by Raoul May — who was described by Kenneth Sisam of Oxford University Press as “a Mexican-American who out of enthusiasm translated it into his best English” — was radically reworked by Paul Harvey. Harvey was a classical scholar who knew no Spanish, but worked by comparing May's manuscript with a copy of the original belonging to the English physiologist Charles Sherrington. The result is as much Harvey's as May's. Later English translations occurred posthumously, and with a few exceptions were from French or German versions of the originals. At least one is an English version of a German translation of a French version of a Spanish original.

Texture of the Nervous System of Man and the Vertebrates: Volume II by Pedro and Tauba Pasik is an English version of Cajal's original *Textura* (for a review of Volume I, see *Nature* **402**, 235–236, 1999). These translators have opted for readability in English, and the result is a workmanlike presentation of the facts assembled by Cajal.

The current volume covers most of the material of volume 2, part 1, of the original, which deals with the brainstem and parts of the diencephalon. The Pasiks have also included text and figures added to the later *Histologie*. This latest translation reveals that the *Histologie* by no means overshadows Cajal's original, as is sometimes suggested. Although the structure of the book was markedly improved in *Histologie* — duplicated numbers for chapters and figures were corrected, marginal comments were added, and typesetting and proofreading errors were eliminated — the basic content is that of the *Textura*. (It is a pity the publishers of the new translation did not emulate the high level of proofreading.)

The literal translation of the word *Textura* in the Spanish title as *Texture* may seem strange to English readers. ‘*Textura*’ conveys the idea of tissue structure, an eighteenth-century concept dating back to the French anatomist Marie-François-Xavier Bichat. Cajal, being well grounded in traditional anatomy, would have been familiar with that usage. ‘*Tissue*’ is probably the more accurate translation, but it risks redundancy, as the nervous system is a tissue in its own right. ‘*Histology*’, as the science of tissues, is probably the best usage, and one that Cajal clearly



An eye for detail: Cajal's schematic representation of the mammalian retina.

favoured in approving the title of Azoulay's translation. It was also used in a facsimile of the *Textura* published in Spain in 1992.

The reproductions of Cajal's figures in that facsimile and in the present translation do not capture the full quality of the illustrations in the original Spanish or French versions. The figures in the present translation have been copied from original drawings in the Cajal Museum, and fine detail has been lost. It is a pity they were not reproduced from the plates prepared for the original printing. For years these remained stacked in a stairwell of the Cajal Institute before being discarded about ten years ago.

Cajal was by no means the universal genius he is often portrayed to have been. His genius lay in his eye for patterns of organization, from which he extracted fundamental biological principles — the basic principles of neuronal organization, circuitry, development and regeneration. But even by the standards of his time, he was relatively unsophisticated in neurophysiology, and on clinical issues he invariably followed the line of contemporary textbooks of neurology.

The world still awaits a disinterested analysis of Cajal's contributions and the extent to which they grew solely out of his genius or rested on general advances in the field. He himself had little doubt that the way to success lay in setting his observations before as wide a scientific readership as possible, and early in his career he submitted French versions of his papers to German- or French-language journals, or persuaded their editors to commission translated versions.

As his reputation grew, he began to be approached by translators. In 1895, Johannes Bressler offered to translate Cajal's works for the library of the publisher J. A.

Barth in Leipzig and eventually translated all the works on the human cerebral cortex. Azoulay, who had some experience with the Golgi technique — which Cajal used to such remarkable effect — wrote to Cajal in 1893 asking if he could translate a series of lectures Cajal had given in Barcelona the previous year on new concepts of nervous organization (*El Nuevo Concepto de la Histología de los Centros Nerviosos*). The lectures defined the intrinsic circuitry of the cerebellum, cerebral cortex, retina and olfactory bulb. This achievement rested on the newly formulated principle of dynamic polarization, which stated that information flow between nerve cells is unidirectional. The lectures eventually appeared as *Les Nouvelles Idées sur la Structure du Système Nerveux chez l'Homme et chez les Vertébrés*.

Cajal would undoubtedly have been delighted to see his works accessible in English, now the predominant language of science. English speakers will detect little of his style in the present translation, but it is hard to capture this without resorting to archaism. Readers seeking to gain some idea of what he could rise to when he let himself go might look into his *Vacation Stories* (University of Illinois Press, 2001), recently translated into English by Laura Otis. The oft-quoted passage from Cajal's autobiography, in which synapse formation between parallel fibre and Purkinje cell is likened to the protoplasmic kiss climaxing an epic love story, pales in comparison with the professor of bacteriology's interpretation of the kymographic tracings that he has surreptitiously obtained during the couplings of his wife and her lover on the laboratory sofa! ■

Edward G. Jones is at the Center for Neuroscience, University of California, 1544 Newton Court, Davis, California 95616, USA.

What are clones?

They're not what you think they are.

Lee M. Silver

“Words that we expect to be generally used ... should be short, euphonious, phonetically spelled, easily pronounced and different from any other word in ordinary use so that it [sic] will not suggest any other meaning than the one desired.” With this explanation, Herbert J. Webber coined the word “clone” in 1903 to describe a colony of organisms derived asexually from a single progenitor. Webber’s contribution to “a more exact expression to our thoughts” found quick acceptance among botanists and gained favour among biologists working with cells in culture. As late as 1965, the science-fiction novel *The Clone* still used the word according to its original connotation (albeit with an absurd plot) to describe a cellular blob expanding across the sewers of a city.

A clone of animal siblings can form naturally, on occasion, as a result of asexual reproduction from a single progenitor embryo. However, in contrast to plants, whole animals cannot be grown directly from cells that have begun to differentiate into a specialized form. Forty years ago, developmental biologists wondered whether this was a true biological limitation or just a technical difficulty. It was this question, rather than interest in cloning *per se*, that motivated John Gurdon to perform the experiments in which he transplanted nuclei from normal frog cells into enucleated eggs to produce adult frogs. Looking back now at Gurdon’s review of his research in a 1968 *Scientific American* article, I am struck that only passing reference is made to clones (as colonies of embryos), with attention focused instead on the developmental potential of differentiated cells.

The popular understanding of cloning has its roots in Alvin Toffler’s 1970 book *Future Shock*. Toffler took a clear scientific concept and muddled it into the fantastical prediction that “man will be able to make biological carbon copies of himself”. Unfortunately, this fictitious version of

cloning was presented in a highly influential, non-fiction book. In one fell swoop, clones morphed from the simple progeny of asexual reproduction to sophisticated products of biological engineering created by scientists bent on controlling nature.

Through the popular media, this version of a clone was rapidly integrated into every major language. Ironically, popularization was helped by the very criteria by which Webber had aimed to ensure proper use of the word. The concept of a clone extended to inanimate objects such as computers (PC clones), as well as becoming a figure of speech to describe people (“Tony Blair is a clone of Bill Clinton”). Until 1997, however, the public felt safe in its knowledge that “real” human clones — biological carbon copies — were still securely in the realm of science fiction.

The sense of security was shattered the day that video clips of Dolly the cloned sheep, prancing in her pen, were beamed down to television screens around the globe. To all appearances, Dolly had been created full-grown within a Frankenstein-like Scottish laboratory. Suddenly, the world had a name and an image to attach to the Promethean-like power of bioengineering.

Insight into the current popular view of cloning can be gleaned from the recent US cloning film, *The 6th Day*. The story begins with a series of newspaper headlines reporting the cloning of Dolly and the completion of the Human Genome Project. The process of cloning is then depicted in stunning detail, including imprinting the DNA as well as the mind and physical features of a living person onto a blank body that is then brought to life.

It is not just the lay public that views cloning in this way. In a recent survey of my students, I discovered that most thought that cloning could do more than just reproduce a genome. A leading US publisher of children’s books recently released *Who Cloned the President?*, presenting the full-grown-copy version of cloning as a scientific fact. The real US president, George W. Bush, has said that “no research to create a human being should take place in the United States”; and Ian Wilmut, who brought us Dolly, writes with Rudolf Jaenisch that “we would never be in favour of [using cloning technology for] copying a person”.

I was recently contacted by a Dutch television producer for my reaction to reports that a fringe religious group was ready to use cloning to bring dead children back to life. For the umpteenth time, I explained that no technology exists for making copies of



Doublespeak: descriptions of Tony Blair as a clone of Bill Clinton show how far the word has strayed from its origins.

AP

people, and that real cloning technology might only lead to the birth of a unique and unpredictable child who had the same DNA sequence as someone else, but nothing more. The producer was abrupt and dismissive: “Dr Silver, you are not aware of what cloning can accomplish. Clones are not what you think they are.”

After years of believing otherwise, I realized that he was right. The scientific community has lost control over Webber’s pleasant-sounding little word. Cloning has a popular connotation that is impossible to dislodge. We must accept that democratic debate on cloning is bereft of any meaning. Science and scientists would be better served by choosing other words to explain advances in developmental biotechnology to the public. ■
Lee M. Silver is in the Department of Molecular Biology and the Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey 08544, USA.

FURTHER READING

Webber, H. J. New horticultural and agricultural terms. *Science* **28**, 501–503 (1903).

Thomas, T. L. & Wilhelm, K. *The Clone* (Berkley, New York, 1965).

Gurdon, J. B. Transplanted nuclei and cell differentiation. *Scientific American* **219**, 24–35 (1968).

Toffler, A. *Future Shock* (Random House, New York, 1970).

Roy, R. & Woodruff, L. *Who Cloned the President?* (Golden Books, New York, 2001).

Weiss, R. in *The Washington Post* 29 March, A10 (2001).
Jaenisch, R. & Wilmut, I. Don’t clone humans. *Science* **291**, 2552 (2001).

Cloning has a popular connotation that is impossible to dislodge.

The Sirens' song

Gerry Melino

Individual cells face three choices: to divide (mitosis), to specialize (differentiate) or to commit suicide (apoptosis). The balance between these ensures tight regulation of cell numbers within organisms. If mitosis proceeded without cell death, an 80-year-old person would have 2 tons of bone marrow and lymph nodes, and a gut 16 km long.

Nevertheless, it took a long time for apoptosis to emerge into the limelight. Apoptosis is over 20 times faster than mitosis, and sightings of dying cells are rare. In contrast to passive cell death (necrosis) — which features leakage and inflammation — apoptotic cells are engulfed and degraded by neighbouring cells without a trace. Various morphologies that we would regard as apoptotic have been observed since the middle of the nineteenth century. But it wasn't until the 1980s that apoptosis was generally credited, when Robert Horvitz and collaborators mapped the fate of every cell in the nematode *Caenorhabditis elegans*, including those that were committed to die. It emerged that cell death is determined by a handful of genes. These 'master switches' have been conserved in evolution so that they, or rather their equivalent families, still orchestrate apoptosis in mammals. However, like life, death is not that simple; the wealth of pro- and anti-apoptotic proteins in cells suggests that suicide is as carefully considered by the cell as it was by Camus.

In stating that it is not possible to enter the same river twice, Heraclitus expressed the irreversibility of time. We, too, undergo continuous changes. What controls the changes in the molecules that form our bodies? What provides our sense of permanence? How do our cells constitute a unified whole?

Gradually, the idea emerged that the stability of the body is maintained by signals that control the life and death of single cells. This is a powerful concept, implying that there are specific survival and death signals, and corresponding receptors on cells. Such social control of life and death are vital in complex multicellular networks such as the immune system and the nervous system, where communication between cells is crucial.

To survive, we must resist many death signals: here the Greek myths provide striking metaphors. To resist the persuasive songs of the Sirens, Odysseus plugged his sailors' ears (blocking the receptors) and tied himself to a mast (blocking signalling). Orpheus, however, resisted the songs by loudly singing and playing his lyre. Thus, he superimposed his life song (anti-apoptotic) over the Sirens' death song (pro-apoptotic). Does social control inevitably imply navigation between conflicting signals?

Social control (altruism, cheating and selfishness) is a powerful evolutionary force, and its signals can be translated to the molecular level. Jacob, Monod and Wyman demonstrated the existence of 'repressors', through which signals interact to modulate gene expression, fine-tuning the activity of enzymes or the binding of ligands and receptors. It is no surprise, therefore, that social control extends to unicellular organisms. Regulated forms of cell suicide also occur in bacteria, in the protozoa *Trypanosoma* and *Tetrahymena* and in the amoeba *Dictyostelium*. In bacteria, several genes, organized as toxin-antidote modules, control the balance between life and death. Most are encoded by plasmids, but some by the bacterial chromosome itself. This is the case for the toxin MazF (which fragments the genome), which is neutralized by the antidote MazE, which in turn is continuously degraded by a protease, ClpP. It seems that, like Sisyphus and the rock he pushes up the mountain — which always rolls down before he reaches the summit — unicellular organisms must continuously inhibit self-destruction.

We can now revisit the idea that social control of cell death in eukaryotes also involves intracellular signals (as Ameisen believes) rather than just signals between cells (as Raff believes). Two billion years ago, the atmosphere changed from being reducing to oxidizing. The reactivity of oxygen and its products allowed the development of complex reactive structures, such as haem centres. This led to more flexible protein structures, more reactivity and more regulation (allostery). Symbioses developed to maximize biological performance — eukaryotic cells seem to be the result of such symbiosis. Some bacteria captured mitochondria, giving rise to proto-

Apoptosis

"There is only one serious philosophical problem. It is suicide. To judge whether life is or is not worth living."
Albert Camus

animal cells; others captured chloroplasts, forming proto-vegetal cells. Thus, a dialectic was established between different elements of the symbiotic partnership. However, the toxic potential of components of the consuming partner, such as the electron-transport chain, necessitate their sequestration within mitochondria; for this reason, mitochondrial damage gives rise to signals that can kill the cell. The cell, in turn, protects itself by producing antidotes such as caspase inhibitors and anti-apoptotic proteins.

Understanding apoptosis may require an appreciation of these ancient symbioses. Any change in equilibrium from the inside (DNA damage, metabolic or cell-cycle aberrations) or the outside (signals and receptors) irreversibly activates suicide. The result is a mitochondrion-centred picture of life and death. But if dialogues exist between other intracellular organelles, this view needs rethinking. What's more, apoptosis seems not to be the sole phenotype of cell suicide. Various pathways of self-destruction coexist in our cells that may have been selected during evolution. And several gene products involved in these death pathways also seem to regulate mitosis and differentiation, blurring the frontiers between 'programmes' of life and death.

With the increased understanding of cell suicide has come a shift in our attitude to many diseases. No longer is it fashionable to think of cancers solely as disorders of mitosis, but rather as a failure of apoptosis. Similarly, diseases such as AIDS, neurodegenerative conditions and auto-immunity may result from too much, rather than too little, apoptosis. The complexity and subtlety of cell death not only allows cells to control their fates, but also offers us new therapeutic ways to control them. The effectiveness and selectivity of these interventions will depend on our capacity to dissect the diverse interplay between molecular mechanisms that regulate mitosis, differentiation and death. ■

Gerry Melino is in the Biochemistry Laboratory, IDI-IRCCS, University Tor Vergata, 00133 Rome, Italy.

FURTHER READING

Ameisen, J. C. *La Sculpture du Vivant* (Seuil, Paris, 1999).
Aravind, L. et al. *Science* **291**, 1279–1284 (2001).
Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. *Trends Genet.* **14**, 410–416 (1998).
Insight: Apoptosis. *Nature* **407**, 769–816 (2000).
Raff, M. C. *Nature* **356**, 397–400 (1992).



Death wish: Odysseus is tempted by the Sirens.

Signs of galactic cannibalism

Amina Helmi

Theories predict that galaxies grow by swallowing smaller galaxies, which may then show up as galactic debris. Astronomers think they have found one such tell-tale structure in a nearby galaxy.

Just as palaeontologists hunt for fossils to reconstruct the evolution of life on Earth, so astronomers hunt for ancient streams of galactic debris to reconstruct the formation history of galaxies. On page 49 of this issue, Ibata and collaborators¹ report the discovery of a giant stream-like structure in the Andromeda galaxy. Andromeda is a large spiral galaxy like the Milky Way, and is one of our nearest and largest galactic neighbours. Ibata *et al.* propose that strong interactions between Andromeda and its smaller companion galaxies led to the formation of this stream of stars. This kind of 'fossil' demonstrates the importance of gravitational interactions in building and shaping mature galaxies, and may further our understanding of the unseen 'dark matter', which constitutes 90% of the mass of the Universe.

Galaxies contain most of the stars in the Universe, yet how they form and evolve remains largely mysterious. Most accepted theories propose that they are the result of mergers and accretion of smaller systems². This idea — the hierarchical build-up of structure in the Universe — has gained substantial support in the past ten years, mostly from observations of distant galaxies caught in the act of forming new systems³. Yet present-day galaxies are expected to contain ancient structures left over from the mergers that led to their formation. This sort of fossil signature should be visible in neighbouring galaxies.

Theories of galaxy formation suggest that spiral galaxies are cannibals, feeding on any smaller systems that fall within their grasp. In the sky, spiral galaxies such as the Milky Way look surprisingly flat, with most of the stars concentrated in the thin galactic disk formed by the spiral arms. Numerical simulations of galaxy mergers predict that large spiral galaxies should also have a spherical 'halo' of diffuse matter stripped from the smaller galaxies they have digested. This galactic halo contains both stars (the stellar halo) and dark matter (the dark halo). When a satellite galaxy is torn apart by a giant galaxy, all that remains of the former satellite are trails of stars in the resulting halo (Fig. 1). This is what Ibata *et al.* were looking for.

Two different approaches are generally used to tackle astronomical problems: statistical analyses of a population of objects, or detailed studies of particular objects, in the hope that they are representative of the

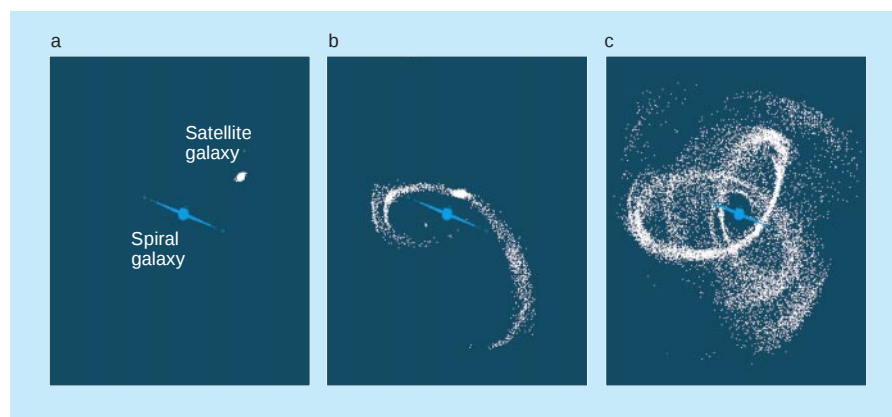


Figure 1 Numerical simulation of a giant galaxy assimilating a smaller companion. a, Initially the stars in the satellite galaxy are densely concentrated. b, The satellite galaxy is progressively destroyed by the gravitational pull of the giant galaxy. c, Eventually the stars from the satellite galaxy populate a roughly spherical halo around the giant galaxy. The halo is much more diffuse than the original satellite galaxy, and contains streams of stellar material as lasting evidence of how it formed.

whole population. In the case of galaxy formation, statistical studies typically focus on the colours or sizes of galaxies and the way they change over time. The complementary approach is to study galaxies that are close enough for individual stars to be imaged and their distances and motions derived. Indeed, such studies of the Milky Way show that there is considerable stellar debris in its halo⁴, supporting the view that our Galaxy has a cannibalistic past⁵.

Ibata and co-workers¹ followed the second approach and carried out a comprehensive survey of stars over a large region in the halo of the Andromeda galaxy. They discovered a puzzling feature in their maps of the stellar distribution — a high-density band of stars in the galactic halo. Their explanation of this feature is that these stars are part of a stream of material stripped from one of Andromeda's companion galaxies (Fig. 2).

Unlike the halo of our Galaxy, whose stars

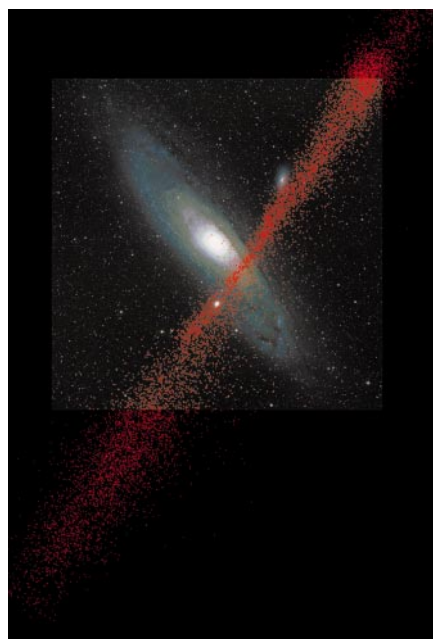


Figure 2 Our nearest neighbour is the Andromeda galaxy, which is a giant spiral galaxy like the Milky Way. Andromeda is visible to the unaided eye, but its spiral structure is difficult to see because the galaxy is viewed almost edge-on. Ibata *et al.*¹ have discovered a stream-like feature in the diffuse halo surrounding the galaxy. The feature is probably the result of interactions between Andromeda and one of its satellite galaxies. The progressive destruction of a dwarf satellite galaxy (such as M32, lower centre, or NGC205, upper right) would add a substantial number of stars to Andromeda's halo. The red stream shows the final distribution of debris from a numerical simulation of the destruction of a satellite galaxy orbiting a spiral galaxy similar to Andromeda. The spatial distribution of the red 'stars' bears a resemblance to the structure observed by Ibata *et al.* Their discovery supports the idea that galaxies are formed through successive mergers of smaller building blocks.

are relatively old and metal-poor, the stars in Andromeda's halo appear to be very metal-rich⁶, as do the stars in the stream-like feature discovered by Ibata and colleagues. Astronomers use the term metal to mean any element heavier than helium, so metallicity is an indication of age because the early Universe was free of heavy elements — these were only produced later, inside stars. The authors propose that Andromeda's entire halo could have formed through the destruction of former companion galaxies, such as the current satellites M32 or NGC205 (the dwarf elliptical galaxies seen in Fig. 2).

This would explain the metal-rich nature of Andromeda's halo, but it still raises the question of why it should be so different from the Milky Way's halo, even though both are similar spiral galaxies. Perhaps the answer lies in understanding the satellite systems surrounding these galaxies. Or perhaps Andromeda and the Milky Way had different formation histories. Alternatively, what is believed to be a metal-rich halo in Andromeda could be an extended thick disk⁷, in which case the stream-like feature observed by Ibata *et al.* could be the result of M32 and NGC205 perturbing this thick disk. There is growing evidence of extreme warps in the main stellar disk of Andromeda, possibly related to the strong perturbations produced by these dwarf galaxies⁸.

To unravel the true nature of the feature described by Ibata *et al.* would require measurements of the stars' spectra to determine their motion, as well as to confirm their metallicity. Knowing how the stars are moving can tell us whether the observed structure is indeed a stream. In that case the stellar motions would also be useful for determining the current orbit of the halo material and its possible history. It remains to be seen whether models can be built to reproduce its characteristics, and in particular whether the dwarf galaxies M32 or NGC205 could be responsible for this unusual stream of material, as the authors suggest.

The sort of detailed analysis of the outer regions of nearby galaxies carried out by Ibata *et al.* complements studies of very distant galaxies. Continuing this work will require detailed images of nearby galaxy halos, as well as analysis of large samples of stars, such as those provided by the Sloan Digital Sky Survey. The analysis of this particular database is presently uncovering the rich structure of our own Galaxy⁴ to an extent unimaginable 15 years ago. In the long run, several satellite missions will accurately measure the motion of thousands to millions of stars in our Galaxy and in our nearest neighbours.

These are exciting times because we are starting to converge on a theory that describes the formation of galaxies from first principles. Observations of the early Universe, of very distant galaxies and now

of our immediate neighbourhood all appear to favour hierarchical formation. Indeed, both Andromeda and the Milky Way — the only two galaxies to be studied in great detail — have revealed a rich substructure in their halos. Today, the main unanswered question is about the nature of the unseen dark matter that keeps satellite galaxies orbiting their parent galaxies, and families of galaxies bound together in a cluster. Studies like those of Ibata *et al.* may be particularly helpful, because streams of debris can be used to map the distribution of both the luminous and the dark matter in a galaxy in unprecedented detail. Questions about the structure of that most elusive of all

galaxy components, the dark halo, may soon have answers. ■

Amina Helmi is at the Max Planck Institut für Astrophysik, Karl Schwarzschildstrasse 1, Postfach 1317, D-85741 Garching bei München, Germany. e-mail: ahelmi@mpa-garching.mpg.de

1. Ibata, R., Irwin, M., Lewis, G., Ferguson, A. M. N. & Tanvir, N. *Nature* **412**, 49–52 (2001).
2. White, S. D. M. & Rees, M. J. *Mon. Not. R. Astron. Soc.* **183**, 341–358 (1978).
3. van den Bergh, S. *et al.* *Astron. J.* **112**, 359–368 (1996).
4. Ivezić, Z. *et al.* *Astron. J.* **120**, 963–977 (2000).
5. Helmi, A., White, S. D. M., de Zeeuw P. T. & Zhao, H. S. *Nature* **402**, 53–55 (1999).
6. Reitzel, D. B., Guhathakurta, P. & Gould, A. *Astron. J.* **116**, 707–722 (1998).
7. Wyse, R. F. G. & Gilmore, G. *Astron. J.* **95**, 1404–1414 (1988).
8. Guhathakurta, P., Choi, P. I. & Reitzel, D. B. *Am. Astron. Soc. Meet.* **197**, 3702 (2000).

Biogeochemistry

The nitrogen fix

James F. Kasting and Janet L. Siefert

At some time in life's history, microorganisms started to make metabolically usable nitrogen from N₂ in the atmosphere. A provocative proposal accounts for the 'why and when' of that event.

To what extent is biological evolution driven by environmental evolution? This is one of the big questions for both evolutionary biologists and geoscientists. On page 61 of this issue, Navarro-González and colleagues¹ describe their investigation into a central innovation in the evolution of life. This was the invention of biological nitrogen fixation, by which life could create biologically usable forms of nitrogen instead of depending on abiotic sources. The authors propose that it was triggered by a change in environmental circumstances — a steep fall in the rate of abiotic nitrogen fixation by lightning at some point during the Archaean era, the first half of Earth's history.

Production of nitric oxide (NO) by lightning is a source of fixed nitrogen today, accounting for 10¹²–10¹³ g of nitrogen per year. But the rate of biological nitrogen fixation is much higher, at least 2 × 10¹³ g N yr⁻¹ and perhaps much more². This process is carried out by certain microorganisms, which ultimately make nitrogen available to all of the rest of life. No organism can do without nitrogen — it is, for instance, a component of all amino acids.

In today's oxygen-rich atmosphere, NO is oxidized to NO₂ and exists in the ocean in the form of nitrate ions, NO₃⁻. Today's atmosphere contains 21% O₂ and 78% N₂. So NO is produced by the efficient (at high temperatures) reaction N₂ + O₂ → 2NO. The Archaean atmosphere is thought to have contained similar amounts of N₂, but little or no O₂. However, NO could still have been produced using oxygen atoms obtained by splitting of CO₂ and H₂O. In the anaerobic Archaean ocean, this NO should have dis-

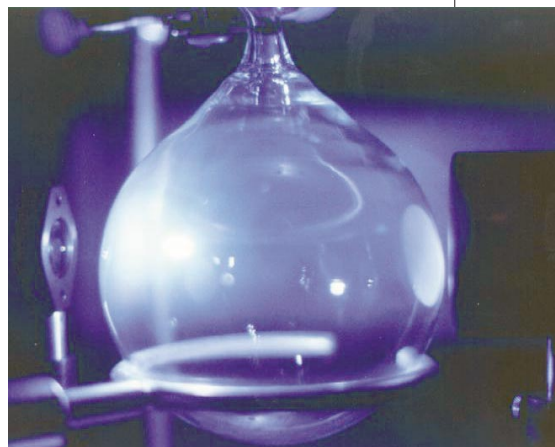


Figure 1 Lightning in the lab — the type of experimental vessel used by Navarro-González *et al.*¹ in their laser simulations of nitrogen fixation in the atmosphere of the early Earth.

proportionated to form N₂O (nitrous oxide), NO₂⁻ (nitrite) and NO₃⁻ (ref. 3). Organisms can use both nitrate and nitrite as a source of nitrogen.

Navarro-González *et al.*¹ have studied the rate of NO production in the laboratory, using lasers to simulate the effect of lightning. They carried out pulsed-laser experiments in reaction vessels (Fig. 1) containing 'atmospheres' with various proportions of N₂, CO₂ and H₂O. This in itself is a useful advance on previous theoretical treatments, most of which assumed that thermodynamic equilibrium is maintained down to some critical 'freeze-out' temperature. In reality, reaction kinetics must be taken into account, as shown by the fact that different results are obtained from mixtures of gases with the same overall

R. NAVARRO-GONZÁLEZ

elemental composition but different distributions of chemical species⁴. In the absence of detailed kinetic information, experiments such as those of Navarro-González *et al.* are clearly a better way to proceed.

The authors' thesis is as follows. On the early Earth, concentrations of CO₂ in the atmosphere were high — because of oxidation of CO produced by impacts of extraterrestrial bodies⁵ and only slow removal of CO₂ by weathering (the continents were smaller at this time⁶, meaning that a smaller area of minerals was exposed for weathering). With these CO₂ conditions, the authors estimate that the initial production rate of NO was about 3×10^{11} g N yr⁻¹. Atmospheric CO₂ levels declined with time, however, as the impact rate dropped and the continents grew. A rise in atmospheric CH₄ produced by methanogenic — methane-generating — bacteria may have warmed the Archaean Earth and speeded the removal of CO₂ by silicate weathering⁷. As this happened, the production rate of NO by lightning dropped to below 3×10^9 g N yr⁻¹ because of the reduced availability of oxygen atoms from the splitting of CO₂ and H₂O. The resulting crisis in the availability of fixed nitrogen for organisms triggered the evolution of biological nitrogen fixation about 2.2 billion years ago.

This hypothesis is attractive but, like many good ideas, difficult to confirm. As the authors point out, genetic studies of the enzyme involved (nitrogenase) indicate that biological nitrogen fixation is an ancient metabolic pathway. If anything, one might imagine that it was invented during the early Archaean, before 3 billion years ago, rather than later on. But then, the crisis in abiotic nitrogen fixation could have occurred very early as well. One model⁸, which emphasizes rapid weathering of fragments from impacts and efficient sequestration of carbon in the Earth's mantle, proposes that concentrations of atmospheric CO₂ were low from the very beginning of Earth's history and that the Archaean climate was warmed almost exclusively by CH₄. In that case, the availability of fixed nitrogen could have been a problem almost as soon as life originated, around 3.5 billion years ago.

The hypothesis also has a couple of wrinkles, the first of which is discussed in the paper. Photochemical models⁹ predict that nitrogen can also be fixed as HCN (hydrogen cyanide) in atmospheres containing N₂ and CH₄. HCN is hydrolysed in solution to form NH₄⁺ (ammonium), which is also a biologically useful form of nitrogen. The rate of HCN production in an anoxic atmosphere is uncertain, but it could have approached the modern rate of abiotic NO production¹. Whether or not this process actually occurred in the postulated manner is unclear, as the relevant photochemical reactions have not been well studied. This deficiency should

be remedied once the importance of the reactions is realized.

The other problem is that HCN might have been involved in a different way. The ancestral counterpart of nitrogenase may have originally been a detoxoyase¹⁰, used to protect organisms against triple-bonded compounds — such as C₂H₂ (acetylene) and HCN — that are formed in anoxic atmospheres. So biological nitrogen fixation may have evolved because too much HCN was being produced, rather than too little. In either case, however, the role of the environment in triggering the evolution of nitrogen-fixing metabolism is clear.

The issues might be clarified by comparing the evolutionary histories of biological nitrogen fixation and methanogenesis. The Archaean climate was probably warm, so the atmosphere must have contained large amounts of either CO₂ or CH₄. Navarro-González and colleagues' argument requires that methanogenesis preceded biological nitrogen fixation. On the face of it, this seems unlikely because the ability to fix nitrogen is widespread among prokaryotes (Bacteria and Archaea), whereas methanogenesis is restricted to a single group within the Archaea.

Parsimony would dictate that the more ubiquitous metabolic process came first. However, if the ability to fix nitrogen was spread by rampant gene transfer between organisms, then methanogenesis could conceivably have evolved first and the premise of Navarro-González *et al.* could hold. This

question might be resolved by comparing the phylogenies of various nitrogenase-related (*nif*) genes with those derived from ribosomal RNA. If the phylogenies of *nif* genes are in accord with those from ribosomal RNA, then nitrogen fixation can be assumed to have been vertically inherited, and so be an anciently derived character. The phylogeny of one component of nitrogenase (*nifH* gene)¹¹ is in partial agreement with that of ribosomal RNA, but still does not constitute conclusive evidence for either vertical or horizontal descent of nitrogen-fixing genes. Analyses of other *nif* genes might allow us to distinguish between these alternatives. ■

James F. Kasting is in the Department of Geosciences, 443 Deike, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

e-mail: kasting@essc.psu.edu

Janet L. Siefert is in the Department of Statistics, Rice University, Houston, Texas 77251-1892, USA.

e-mail: siefert@stat.rice.edu

1. Navarro-González, R., McKay, C. P. & Mvondo, D. N. *Nature* **412**, 61–64 (2001).
2. Falkowski, P. G. *Nature* **387**, 272–275 (1997).
3. Mancinelli, R. L. & McKay, C. P. *Origins Life* **18**, 311–325 (1988).
4. McKay, C. P. & Borucki, W. J. *Science* **276**, 390–392 (1997).
5. Kasting, J. F. *Origins Life* **20**, 199–231 (1990).
6. Walker, J. C. G. *Origins Life* **16**, 117–127 (1985).
7. Pavlov, A. A., Kasting, J. F., Brown, L. L., Rages, K. A. & Freedman, R. J. *Geophys. Res.* **105**, 11981–11990 (2000).
8. Sleep, N. H. & Zahnle, K. J. *Geophys. Res.* **106**, 1373–1399 (2001).
9. Zahnle, K. J. *J. Geophys. Res.* **91**, 2819–2834 (1986).
10. Fani, R., Gallo, R. & Lio, P. J. *Mol. Evol.* **51**, 1–11 (2000).
11. Braun, S. T., Proctor, L. M., Zani, S., Mellon, M. T. & Zehr, J. P. *FEMS Microbiol. Ecol.* **28**, 273–279 (1999).

Apoptosis

DNA destroyers

Michael O. Hengartner

Proteins with quite mundane functions in healthy cells often behave very differently during cell suicide. One protein normally involved in copying mitochondrial DNA actually degrades nuclear DNA in dying cells.

The degradation of DNA is one of the hallmarks of programmed cell death (apoptosis)¹. When forced to commit suicide, apoptotic cells — like good secret agents — grimly destroy their 'instruction book', chewing up their genomic DNA into tiny morsels. Until now, only two DNA-destroying enzymes (nucleases) with a clear role in cell death were known, one in mammals² and one in the nematode worm *Caenorhabditis elegans*³. But, on pages 90–99 of this issue, Li and colleagues⁴ and Parrish and co-workers⁵ show that another nuclease, endonuclease G (endoG), also contributes to the carnage, and might even influence the likelihood that a cell will live or die.

Li *et al.*⁴ encountered endoG while searching for new cell-death-inducing factors that are released from mitochondria — the energy-generating organelles — in

mammalian cells that are primed to die. The authors exposed isolated mitochondria to the truncated, apoptosis-inducing version of the protein Bid. The mitochondria then released a nuclease activity into the medium; this activity could, when incubated with isolated nuclei, generate the pattern of DNA fragments that is characteristic of apoptotic cells. The nuclease activity was distinct from that of the known mammalian nuclease, CAD (also known as DFF40)². Li *et al.* purified the new apoptotic nuclease, and identified it as endoG — a previously described mitochondrial enzyme with a proposed role in the replication of mitochondrial DNA⁶.

Why would a cell need more than one nuclease to digest its DNA? One hint might come from the different ways in which CAD and endoG are regulated. In living cells, CAD is sequestered in an inactive complex, bound

to its inhibitory subunit (known as ICAD or DFF45). In response to death-inducing stimuli, cells activate a chain of enzymes called caspases. One of these caspases cleaves ICAD, releasing the active nuclease, which is then free to wreak havoc in the dying cell's nucleus². By contrast, endoG is activated by a change in its subcellular localization. Normally, endoG is found only in mitochondria, but in response to apoptotic stimuli it is released into the body of the cell⁴. Once released, endoG — like apoptosis-inducing factor⁷, another mitochondrial protein required for cell death — will work even when caspase activation is limited or compromised, as might be the case, for example, during viral infection.

It is also possible that different apoptotic nucleases have distinct functions. Indeed, Parrish *et al.*⁵ suggest that endoG not only participates in the 'deconstruction' of apoptotic cells, but can also contribute to the actual killing process. Using a genetic screen in *C. elegans*, these authors identified genes that, when mutated, protect cells from an activated form of the caspase CED-3. One of the genes identified, *cps-6*, probably encodes the nematode counterpart of endoG — the CPS-6 protein is also found in mitochondria, and shows nuclease activity towards both isolated nuclei and purified plasmids (small, circular DNA molecules). Indirect evidence indicates that CPS-6 probably also participates in DNA degradation in apoptotic cells in living worms⁵.

Surprisingly, unlike the two previously known apoptotic nucleases^{2,3}, CPS-6 not only promotes DNA degradation, but also appears to contribute to cell killing. Loss of *cps-6* function slightly increased cell survival in many of the genetic backgrounds tested by the authors, including worms in which CED-3 was either weakly active, or even — rather surprisingly — completely inactive⁵. The implication is that endoG is not only sufficient to cause cell death, but might also be necessary, at least under certain conditions.

Like any good story, these two papers^{4,5} leave us thirsting for more. First, how do we reconcile the previously described function for endoG — in the replication of mitochondrial DNA — with this new role in apoptosis? To the experienced apptologist, this is hardly an issue. Ever since the discovery of the double life of the cytochrome *c* protein, as both an electron carrier and a pro-death agent, we have grown accustomed to the idea that apoptotic proteins might have quite mundane functions in normal cells.

A more serious problem arises, however, when one considers the expected localization of endoG within mitochondria. Unlike cytochrome *c*, which is found in the space between the two membranes of the mitochondrion, endoG should be within the

inner matrix, deep inside the mitochondrion, if it is to participate in mitochondrial DNA replication. Yet we know that apoptotic mitochondria retain most matrix proteins. Indeed, Li *et al.*⁴ show that the matrix protein Hsp70 is not released upon treatment with truncated Bid. The authors suggest that endoG might actually be found mostly in the intermembrane space, with only a small fraction participating in mitochondrial DNA replication in the matrix. Detailed organelle-fractionation analyses should settle this question.

The finding that endoG can promote apoptosis in *C. elegans* is important: it is the strongest evidence yet that mitochondria help to regulate cell death in nematodes. (Although the worm counterpart of the mammalian anti-apoptotic protein Bcl-2 binds to mitochondria⁸, there is as yet no evidence that cytochrome *c* is involved in nematode apoptosis.) Unfortunately, Parrish *et al.*⁵ do not provide any direct evidence that endoG leaves the mitochondria in *C. elegans*, as Li *et al.* do for mammalian endoG. However, because the loss of the worm endoG alters the kinetics of DNA degradation in apoptotic nuclei, we have to assume that it does leave the mitochondria. How it does this is a mystery, as *C. elegans* apparently lacks counterparts of Bax and Bak, mammalian proteins that are thought to be essential for allowing mitochondrial molecules to escape⁹.

It is surprising that the loss of endoG activity in *C. elegans* results in increased cell survival. Studies of the previously known apoptotic nucleases^{2,3} had suggested that DNA degradation has a late, non-essential

role in apoptosis. Perhaps the nematode endoG acts earlier in the apoptotic pathway, at a point when a cell's fate has not yet been irreversibly sealed. In cells wavering between life and death, removal of even a single apoptotic subprogramme might be sufficient to tip the balance towards survival. Indeed, experiments in my laboratory show that eliminating the subprogramme by which dying cells are engulfed by other cells significantly increases cell survival when caspases are activated only weakly. Whether the mammalian endoG also shows cell-killing activity remains to be determined.

Finally, it is worth pointing out that endoG is not restricted to multicellular organisms — a fact that is perhaps not surprising, given its proposed role in mitochondrial DNA replication. It will be interesting to see where on the evolutionary tree endoG acquired its death-promoting character. Perhaps this nuclease is responsible for some of the caspase-independent cell deaths and ladders of fragmented DNA that have been observed in plants, fungi and protozoa. ■

Michael O. Hengartner is at the Institute of Molecular Biology, University of Zurich, Winterthurerstrasse 190, Zurich CH-8057, Switzerland.

e-mail: michael.hengartner@molbio.unizh.ch

1. Wyllie, A. H. *Nature* **284**, 555–556 (1980).
2. Nagata, S. *Exp. Cell Res.* **256**, 12–18 (2000).
3. Wu, Y. C., Stanfield, G. M. & Horvitz, H. R. *Genes Dev.* **14**, 536–548 (2000).
4. Li, L. Y., Luo, X. & Wang, X. *Nature* **412**, 95–99 (2001).
5. Parrish, J. *et al.* *Nature* **412**, 90–94 (2001).
6. Cote, J. & Ruiz-Carillo, A. *Science* **261**, 765–769 (1993).
7. Lorenzo, H. K., Susin, S. A., Penninger, J. & Kroemer, G. *Cell Death Differ.* **6**, 516–524 (1999).
8. Chen, F. *et al.* *Science* **287**, 1485–1489 (2000).
9. Hengartner, M. O. *Nature* **407**, 770–776 (2000).

High-energy physics

Neutrinos reveal split personalities

John N. Bahcall

For more than 30 years scientists have puzzled over the mystery of the missing neutrinos emitted from the Sun. Data from underground detectors in Canada and Japan combine to provide the answer.

Neutrinos are unique subatomic particles. They have no electrical charge, travel essentially at the speed of light, and come in three types: electron, muon and tau. These particles are so elusive that you do not notice the hundred billion solar neutrinos that pass through your thumbnail every second. The 'solar neutrino mystery' began in 1968, when a pioneering experiment¹ found fewer electron-type solar neutrinos than predicted by a detailed model² of how the Sun shines (and how many neutrinos are detected). Two ideas were widely discussed: either the model of the Sun was wrong, or something happens to the neutrinos on their way to the Earth.

In the 1980s and 1990s, several international teams of scientists performed a range of underground experiments designed to solve this mystery. This work provided circumstantial evidence that some solar neutrinos change en route to the Earth from the electron-type produced in the Sun to another type that is harder to detect. But there was no smoking-gun evidence for such particle personality changes, usually called neutrino oscillations. This situation changed dramatically three weeks ago when scientists working in Canada announced that they had solved the mystery³. "The Sudbury Neutrino Observatory (SNO) finds that the solution lies not with the Sun, but



100 YEARS AGO

Catalase is the name given to a new enzyme of general occurrence described by Dr. Oscar Loew in Report (No. 68) of the U.S. Department of Agriculture (Division of Vegetable Physiology and Pathology) with special reference to the tobacco plant. This enzyme possesses the power of producing catalytic decomposition of hydrogen peroxide, a decomposition which, according to the author's experiments, is probably not produced by any other known enzyme. The enzyme appears to exist in an insoluble and in a soluble form, which are designated α - and β -catalase respectively... Experiments on the nature of catalase indicate that it is an oxidising enzyme, the most characteristic reaction studied in this direction being its rapid oxidation of hydroquinone to quinone. Numerous tests have established the general occurrence of catalase in the vegetable kingdom. No living plant or vegetable organ tested was found free from it, some plants containing more of the soluble, others more of the insoluble, form. In the animal kingdom it also appears to be widely distributed. From *Nature* 4 July 1901.

50 YEARS AGO

A situation provoking speculation has been found during the course of a preliminary investigation of one-shear ($1\frac{1}{2}$ -year old) rams offered for sale at the Feilding Stud Ram Fair... Data used in these analyses were derived from the catalogues of the 1948, 1949 and 1950 sales and the 1948 Flock Book of the New Zealand Romney Marsh Breed Society. Of the 612 rams offered for sale in the three years, 52.6 per cent were sired either by lambs or by one-shear rams. The offspring of the one-shear rams brought the highest average price and had a lower rejection-rate than any other age... However, Goot has shown that one-shear sires comprise only 27.8 per cent of the sires available for use in those flocks static in numbers and consisting of 400 or more ewes. Selection of rams for the Stud Fair and for single-entry in the Flock Book is almost solely on phenotype and certainly without reference to age of the parents, nor does this latter point interest buyers. Because of this, it would be expected that the sires of rams chosen for single-entry and for sale would be represented in the same proportion as they are used in the flocks... Further analyses are planned to throw more light on these rather perplexing problems. From *Nature* 7 July 1951.

with the neutrinos, which change as they travel from the core of the Sun to the Earth."

How did the SNO scientists — a collaboration of 113 scientists from 11 universities and laboratories in Canada, the United States and the United Kingdom — solve the solar neutrino mystery? Over 2,000 metres below the Earth's surface, within an active copper and nickel mine, the SNO collaboration built a laboratory⁴ the size of a ten-storey building. Here, their underground detector is shielded from cosmic rays and radioactive contamination from dust — the laboratory is so clean it contains less than a teaspoon of dust. SNO scientists built a spherical detector, 12 metres in diameter, which contains 1,000 tonnes of heavy water (D_2O) and is itself immersed in a 30-metre cavity filled with normal water (H_2O). Neutrinos from the Sun are occasionally detected by the heavy water (about five per day), producing light that is measured by 10,000 photomultipliers.

In the initial results reported by SNO, only electron neutrinos were detected (by a specific reaction in the heavy water). A Japanese-American experiment, known as Super-Kamiokande⁵, can detect all three types of neutrinos, but is mostly sensitive to electron neutrinos. But Super-Kamiokande, which uses pure H_2O in an underground detector in northern Japan, does not distinguish between electron-type and other solar neutrinos.

If only electron neutrinos travel from the Sun to the Earth, then SNO and Super-Kamiokande would measure the same number of neutrinos. If some solar neutrinos are muon or tau neutrinos, then Super-Kamiokande would measure a larger number. Indeed, the Super-Kamiokande number exceeds the SNO number with a probability of 99.96% (3.3 standard deviations), conservatively calculated. This is a smoking gun.

The Sun emits neutrinos over a wide range of energies, but SNO and Super-Kamiokande are sensitive to a specific energy range. Using data from both measurements, SNO scientists calculated the total number of these solar neutrinos that reach the Earth. The measured number agrees well (within 0.3 standard deviations) with the prediction of the standard solar model⁶.

What does all this mean? In 1969 two Russian scientists first proposed⁷ that neutrino oscillations cause the observed discrepancy between the predicted and measured numbers of solar neutrinos. In 1998, experiments at the Super-Kamiokande detector⁸ provided the first evidence of neutrino oscillations by studying neutrinos produced when cosmic rays interact with the Earth's atmosphere. SNO has now confirmed that solar neutrinos undergo oscillation.

The Sun only produces electron neutrinos, but some muon or tau neutrinos reach us from the Sun. Therefore, solar neutrinos

must oscillate from one type to another. This phenomenon, which requires that neutrinos have non-zero mass, is not predicted by the simplest version of the textbook theory of weak particle interactions (called electroweak theory). The standard theory of weak interactions must be modified slightly, which is important but not unexpected. Most importantly, the specific way neutrinos oscillate, which must be determined by future experiments, may help select the correct generalization of existing physical theories.

Neutrinos contribute to the mass density of the Universe. Combining results on neutrino masses from SNO, Super-Kamiokande and nuclear physics experiments, SNO scientists conclude that electron, muon and tau neutrinos contribute between 0.1% and 18% of the critical mass density of the Universe. The most plausible value is 0.1%. A neutrino mass density of 0.1% is probably too small to affect significantly the geometry or fate of the Universe, but it is about one-quarter of the mass density of all the stars we observe. So even though there is an enormous number of neutrinos in the Universe, the small amount of mass they contribute is not going to solve the problem of the Universe's missing 'dark matter'.

Arguably, the most spectacular result from SNO is that the total number of solar neutrinos measured by the observatory and Super-Kamiokande is bang on that predicted by the standard solar model. In appropriate units, the predicted value is 5.05 ± 0.2 and the measured value inferred by comparing the results of SNO and Super-Kamiokande is 5.44 ± 1.0 . This is a triumph for the theory of stellar evolution. The predicted number of neutrinos depends on the 25th power of the central temperature of the Sun. Getting the neutrinos correct to 20% implies we can calculate the Sun's central temperature (15.7 million kelvin) to better than 1%. As stellar evolution theory is widely used to interpret observations of stars and galaxies, this agreement is a cause for rejoicing among astronomers.

Physicists are happy because they have an interesting phenomenon to study; astronomers are happy because their solar theory has been proven correct. But the work has only just begun. Scientists have so far made detailed measurements of only 0.005% of the neutrinos astronomers believe are emitted by the Sun. The remaining neutrinos are at lower energies and so are more difficult to detect. Until these lower-energy neutrinos are observed and compared with theory, we cannot be sure we really understand the intricacies of the mystery of the missing neutrinos. In the meantime, SNO and Super-Kamiokande have crucial additional measurements to make. It is a great time to be involved in neutrino research. ■

John N. Bahcall is at the School of Natural Science,

Institute for Advanced Study, Princeton, New Jersey 08540, USA.

<http://www.sns.ias.edu/~jnb>

1. Davis, R. Jr, Harmer, D. S. & Hoffman, K. C. *Phys. Rev. Lett.* **20**, 1205–1209 (1968).
2. Bahcall, J. N., Bahcall, N. A. & Shaviv, G. *Phys. Rev. Lett.* **20**, 1209–1212 (1968).
3. Ahmad, Q. R. *et al.* <http://arXiv.org/abs/nucl-ex/0106015>

4. McDonald, A. B. *et al.* *Nucl. Phys. Proc. Suppl.* **91**, 21–28 (2000).
5. Fukuda, S. *et al.* *Phys. Rev. Lett.* (in the press), <http://arXiv.org/abs/hep-ex/0103032>
6. Bahcall, J. N., Pinsonneault, M. & Basu, S. *Astrophys. J.* **55**, 990–1012 (2001), <http://arXiv.org/abs/astro-ph/0010346>
7. Gribov, V. N. & Pontecorvo, B. M. *Phys. Lett. B* **28**, 493–496 (1969).
8. Fukuda, Y. *et al.* *Phys. Rev. Lett.* **81**, 1562–1567 (1998).

Evolutionary biology

Searching for speciation genes

Roger Butlin and Michael G. Ritchie

The formation of new animal species often results from divergence in male sexual behaviours and female preferences. The genetic basis of this sexual isolation in fruitflies is gradually being revealed.

What are the genetic changes that underlie the origin of new species? Are there common patterns across taxonomic groups, and what do these patterns tell us about the evolutionary circumstances that promote speciation? Writing in *Proceedings of the National Academy of Sciences*, Ting *et al.*¹ and Doi *et al.*² arrive at contrasting conclusions. Both groups studied fruitflies, and asked how many genes need to be changed before a female switches her mating preference from males of one species to males of another. Depending on the pairs of species studied, the answer may be either many¹ or very few — perhaps only one².

Research on the genetics of speciation began in earnest with Dobzhansky³ in the 1930s, and those looking for 'speciation genes' have used similar techniques ever since. To map the chromosomal locations (loci) of such genes, researchers use pairs of related species that do not normally mate but will do so in the laboratory if given no other choice. The hybrid offspring are then examined for traits that are related to 'reproductive isolation' between the two original species. Such traits might include sterility of the hybrids (a cause of postmating isolation) or differences in mate preference (which contributes to premating isolation). At the same time, researchers look for 'marker' genes that are associated with the isolating traits⁴. This provides clues to the location of the genes that underlie reproductive isolation between the original species; the assumption is that the divergence of these genes and traits contributed to the divergence of the species. But the problem with identifying such speciation genes is that reproductive isolation is usually a by-product, not their main function, and arises only when two species interact.

Most progress has been made on the genetics of postmating isolation, specifically the sterility of hybrid males. When new forms of genes (alleles) arise in diverging populations, they clearly do not spread because they cause sterility; they must be

'fixed' for other reasons. Mapping these loci has borne out Dobzhansky's suggestion³ that hybrid sterility is based on harmful interactions between incompatible alleles that have been fixed independently in diverging populations⁵. A common pattern is that many loci are involved⁶; identification of the genes at these sites reveals the main functions of the loci, and allows one to assess whether divergence was promoted by natural or sexual selection, or by random 'genetic drift'⁷.

Sexual (prematuring) isolation may be the main cause of speciation in animals. It is often brought about by divergence in male sexual signals and female preferences. Here, the key genetic issues are different. The main trait to look for is assortative mating, in

which individuals prefer to mate with individuals who resemble themselves. But this is difficult to study. It involves up to four types of individual (males and females of both species); it is hard to separate from variation in readiness to mate; and it is based on flexible behavioural decisions that are often context-dependent and difficult to quantify. If, on the other hand, one studies the genetics of male signals or female preferences, it is hard to know how they relate to isolation. So the genetics of sexual isolation has not made as much progress as the analysis of sterility, and little consensus has emerged⁸.

This is where the new papers^{1,2} come in. Ting *et al.*¹ have studied the sexual isolation of the widespread M form of *Drosophila melanogaster* and the Z form found in southern Africa. Females of the Z form strongly prefer Z males over M males⁹. The authors started with flies that had an M genome, except that the two copies of chromosome III were replaced by the Z form of these chromosomes. From here, Ting *et al.* produced flies in which sections of either one or both of the Z chromosomes were replaced by sections from the corresponding M chromosome. Tests of assortative mating using females from these lines and males from reference stocks allowed the authors to map regions of chromosome III that influence female traits (Fig. 1). The reciprocal crosses allowed male traits to be mapped. The results indicate that at least four loci on chromosome III influence male behaviour and at least three affect female

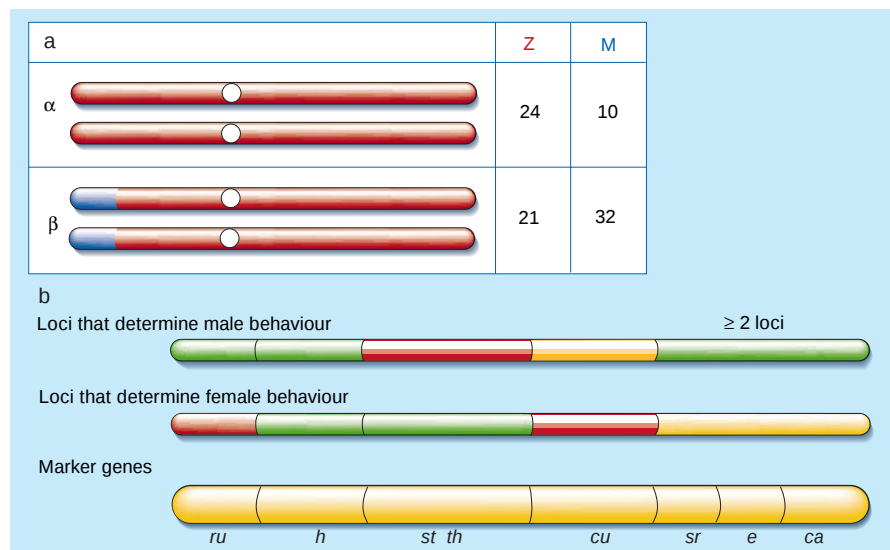


Figure 1 How many genes cause speciation? Ting *et al.*¹ tested assortative mating — an individual's choice of mating partner — between pairs of fruitflies (*Drosophila melanogaster*). The reference stocks were the Z and M forms of this species. The 'substitution' stocks (α and β) had a genetic background that was broadly that of the M forms, but differed from the reference lines in portions of either one or both copies of chromosome III. **a**, In this case, in the β line, both copies of chromosome III have a small region of M genes (blue). These β females mate significantly more with M males than do α -line females, in which the whole of each chromosome III is of the Z form (red). The numbers represent numbers of mating pairs. **b**, Combining results across pairs of different lines reveals regions on chromosome III that have significant effects alone (red) and pairs of regions that influence mating together but not separately (green).

behaviour. The authors conclude that pre-mating isolation has a diffuse genetic basis.

Doi *et al.*², meanwhile, studied isolation between the cosmopolitan *D. ananassae* (A) and its Melanesian sibling species *D. pallidosa* (P). These species are almost completely sexually isolated. But the authors show that A females no longer discriminate strongly against P males if the males are prevented from singing their 'love songs' (by removal of their wings) or if females are prevented from hearing them (by removal of their 'ears'). This suggests that divergence in male song patterns and associated female preferences underlies the sexual isolation (and speciation) in this case.

First-generation female hybrid offspring of A and P flies showed the same pattern of discrimination as A females, implying that the A genes involved are dominant over the P genes. Most of these genes mapped to chromosome II and, remarkably, repeated backcrossing revealed that only a single chromosomal region, marked by the *Delta* gene, was required for hybrids to prefer A males. So, a small locus — perhaps even a single dominant allele — might underlie the preference of A females for A males. Surprisingly, this region did not affect the readiness of A females to mate with P males. Perhaps the preferences for A and P males have a different genetic basis, or perhaps the gene (or genes) in the region marked by *Delta* controls willingness to mate rather than preference.

There are some caveats to the single-gene interpretation. The *Delta*-containing region might be larger (and so include more genes) than the authors suspect; characterization of more molecular markers in this region will help here. Nonetheless, if a single gene could be identified and cloned, it would be a significant addition to the poor list of speciation genes. At the moment, we really have no idea what a 'preference' gene might be like. Another question is whether genes that influence male song map to the same region.

The conclusions of the two studies^{1,2} are strikingly different — a fact that may be related to variations in behavioural complexity. It seems that the divergence of acoustic signals alone explains the isolation between the A and P species², whereas the behavioural basis of mate choice in the M and Z forms involves many types of signal¹⁰. The differences in experimental design may have accentuated this effect: Doi *et al.* measured the likelihood of A females accepting A or P males, but did not offer a choice, whereas Ting *et al.* measured relative mating success in a multiple-choice design.

Alternatively, the different genetic architectures may reflect the histories of these populations. The M and Z forms seem to have diverged while in the same areas, but the A and P species may have evolved while geographically separate. Ting *et al.* suggest that the M and Z forms of *D. melanogaster*

might not be able to evolve to the point at which they are separate species because too many genes are involved in sexual isolation. With such different patterns apparent in the few studies available, many more systems will have to be analysed before a really informative picture emerges.

Roger Butlin is at the School of Biology, University of Leeds, Leeds LS2 9JT, UK.
e-mail: r.k.butlin@leeds.ac.uk

Michael G. Ritchie is at the School of Biology, University of St Andrews, St Andrews KY16 9TS, UK.

e-mail: mgr@st-and.ac.uk

Quantum physics

Air juggling and other tricks

Eric J. Heller

Quantum tunnelling breaks the rules of classical physics — and leads to ghost-like transfer of matter through barriers. Demonstrations of a new type of quantum tunnelling have the ghosts taking new liberties.

An old joke goes like this: a motorist stops to ask a farmer how to get to a village a few miles away. After much thought, the farmer says with conviction: "You can't get there from here."

The farmer may have been a classical physicist. Given constraints such as energy conservation, certain types of motion are isolated in classical systems — one type never leads to the other. Usually we think of a hill or energy barrier preventing the journey, but often the barriers are more subtle and indirect. Imagine a ball bouncing between two semicircular mirrors (Fig. 1). Classical dynamics forever confines it to the region between the mirrors, even though there is no energy barrier preventing it leaving through one of the open gaps. Classical motion does not allow escape, but quantum mechanics is famous for allowing tunnelling into classically forbidden barriers, and even right through them.

Can quantum systems wriggle out of subtler, dynamical barriers like those presented by the two mirrors? Absolutely. Delicate experiments in the Phillips laboratory¹ (described on page 52 of this issue), and in the Raizen laboratory² (published by *Science*), demonstrate 'dynamical tunnelling'^{3–5} of ultracold atoms, which allows them to transfer between two stable, but classically separate, states of motion.

Dynamical tunnelling is a close cousin of 'above-barrier reflection', in which a particle with enough energy to go over a barrier is nonetheless reflected back — an event forbidden by classical physics. This is also sometimes called diffraction, but this term is used in so many contexts (some of them classically forbidden, some not) that it is best avoided. Dynamical tunnelling can have remarkable

1. Ting, C.-T., Takahashi, A. & Wu, C.-I. *Proc. Natl Acad. Sci. USA* **98**, 6709–6713 (2001).
2. Doi, M., Matsuda, M., Tomaru, M., Matsubayashi, H. & Oguma, Y. *Proc. Natl Acad. Sci. USA* **98**, 6714–6719 (2001).
3. Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
4. Tan, C. C. *Genetics* **31**, 558–573 (1946).
5. Wu, C.-I. & Palopoli, M. F. *Annu. Rev. Genet.* **28**, 283–308 (1994).
6. True, J. R., Weir, B. S. & Laurie, C. C. *Genetics* **142**, 819–837 (1996).
7. Ting, C.-T., Tsaur, S. C., Wu, M. L. & Wu, C.-I. *Science* **282**, 1501–1504 (1998).
8. Ritchie, M. G. & Phillips, S. D. F. in *Endless Forms: Species and Speciation* (eds Howard, D. A. & Berlocher, S.) 291–308 (Oxford Univ. Press, Oxford, 1998).
9. Hollocher, H. *et al.* *Evolution* **51**, 1175–1181 (1997).
10. Colegrave, N., Hollocher, H., Hinton, K. & Ritchie, M. G. *J. Evol. Biol.* **13**, 143–150 (2000).

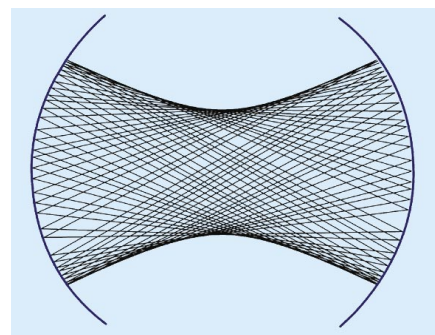


Figure 1 Classical and quantum motion of a bouncing ball. Two semicircular mirrors confine a bouncing ball to a regular pattern of motion. Even though escape is possible energetically, and no potential barrier would prevent it, the classical motion of the ball does not allow it to escape. Quantum mechanically, it would be able to slowly escape, by 'dynamical tunnelling'. This sort of quantum tunnelling has been discussed theoretically, and has now been observed directly in experiments^{1,2} with ultracold atoms trapped in laser fields.

and non-intuitive consequences. Consider the formaldehyde molecule, H_2CO , spinning about the C–O axis with the oxygen atom pointing up (Fig. 2, overleaf). Classically, the oxygen is doomed to point up forever, but in quantum mechanics it can oscillate between pointing up and down. It does this without violating the conservation of energy or angular momentum. In reality, a rotating formaldehyde molecule that starts with oxygen pointing up does flip its direction, just as predicted by quantum mechanics.

By using relatively large numbers of atoms, the Phillips and Raizen groups^{1,2} have caught extremely cold quantum gases in the act of doing something impossible for

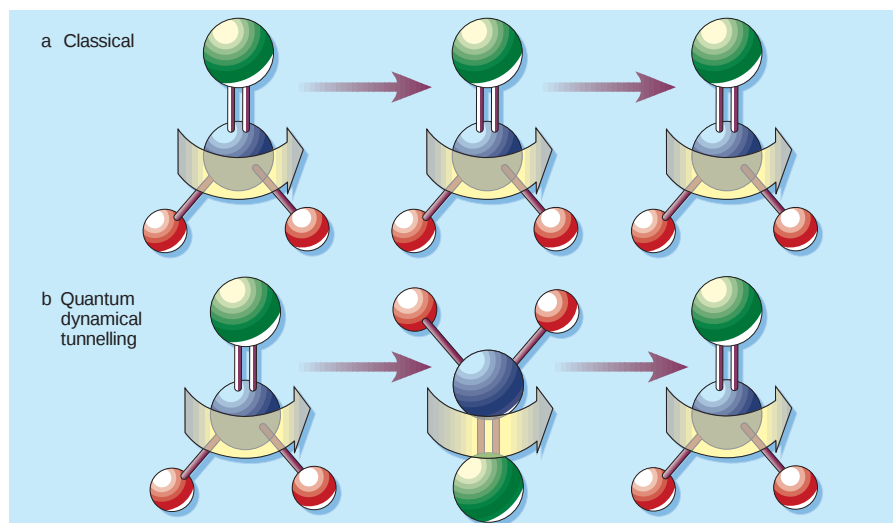


Figure 2 Classical and quantum motion of a formaldehyde molecule. **a.** In the classical picture, a rotating H_2CO molecule always stays in the same orientation, with the oxygen atom pointing upwards. **b.** In the quantum picture, the H_2CO molecule can flip between an oxygen-up and an oxygen-down state through a process known as dynamical tunnelling.

classical particles. The atoms were put into a very distinct kind of motion, but were later seen in the wrong place at the wrong time (if they had continued to behave classically). Specifically, they were caught travelling in the wrong direction — a feat that is possible only if they had used dynamical tunnelling to get there. Classical particles would need a specific kick to change their direction.

The two groups of experimentalists used a web of crossed laser beams to create elaborate three-dimensional force fields in which the intensity of the light varies periodically. This sort of 'optical lattice' was first created in the early 1990s. When ultracold atoms are added to the lattice they are attracted or repelled from regions of strong laser intensity, depending on the colour (frequency) of the laser beams, which are kept far from an atomic absorption frequency. By varying the strength of the laser light, the experimentalists can control the positions and motions of the atoms. The result is like a juggling act, in which the balls (atoms) are kept in motion in space by precise forces exerted at just the right time.

But the juggling acts performed by the Phillips¹ and Raizen² groups have a twist. Imagine an identical juggler standing next to the first one. He is 'air juggling' — that is, he has nothing to juggle with and is just going through the motions. The first juggler does not throw his balls to him, but even so the second juggler finds that after a time he has the balls, and the first becomes the air juggler. And then the first juggler has the balls again, and so on. This is dynamical tunnelling. The Raizen group achieved it with thousands of atoms, and the Phillips group with millions of atoms in a Bose–Einstein condensate, a form of matter in which all the atoms have the same quantum state.

But what is happening at that magical

halfway point, when the balls have not completely tunnelled from one juggler to the other? At this point, the balls are in both places at once with equal probability — a feature known as quantum coherent superposition, and an essential ingredient of any approach to building quantum computers, for example. So the demonstration of dynamical tunnelling is also a demonstration of quantum coherent superposition of distinct events — all of the atoms were travelling in both directions at once. This is a fact of life in the quantum realm.

Both experimental groups worked with systems containing a degree of chaotic motion, which makes things more challenging theoretically. They did not do this deliberately — the moving optical field they created with the laser beams induces regions of classical chaos. But it raised the possibility that the process leading to the atoms going the wrong way was classical chaotic motion, rather than quantum tunnelling. Chaos is an aspect of classical systems that corresponds to extreme sensitivity to initial conditions, and often

leads to rapid, seemingly random cycling between different kinds of motion. At the suggestion of Vitali Averbukh of the Technion in Israel, the Phillips group took pains to rule out the possibility that classical chaotic transport was heavily involved, thereby confirming that dynamical tunnelling was taking place.

These experiments also raise the possibility of an even newer tunnelling concept — chaos-assisted tunnelling⁶. Chaos can co-exist with regions of stable, non-chaotic motion because some types of motion, called regular motion, can avoid getting mixed up in the chaotic fray. In this regime, chaos can assist tunnelling by providing a 'free ride' over to another zone of regular motion once the system has tunnelled out of the first regular zone into the chaotic region.

Many previous experiments have demonstrated quantum tunnelling by individual atoms or molecules, but a nearly macroscopic system containing millions of atoms might be expected to behave more classically. Certainly near-macroscopic tunnelling has been seen before, as in the Josephson effect in superconductors or in barrier tunnelling by Bose–Einstein condensates⁷, but such observations are rare, and physicists are always hungry for more examples. From a broader perspective, these and other recent experiments demonstrate that it is possible to exert quantum control over ultracold atoms with astonishing finesse and coherence. We can look forward to a continuing stream of mind-bending examples, perhaps leading to a better understanding of the implications of quantum mechanics. ■

Eric J. Heller is in the Departments of Physics and Chemistry, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: heller@physics.harvard.edu

1. Hensinger, W. K. *et al.* *Nature* **412**, 52–55 (2001).
2. Steck, D. A., Raizen, M. G. & Oskay, W. H. *Science* **5 July** 2001 (10.1126/science.1061569).
3. Davis, M. J. & Heller, E. J. *J. Chem. Phys.* **75**, 246 (1981).
4. Davis, M. J. & Heller, E. J. *J. Phys. Chem.* **85**, 307 (1981).
5. Lawton, R. T. & Child, M. S. *Mol. Phys.* **44**, 709–723 (1981).
6. Tomsovic, S. & Ullmo, D. *Phys. Rev. E* **57**, 1421 (1998).
7. Anderson, B. P. & Kasevich, M. A. *Science* **282**, 1686–1689 (1998).

Ecology

Price put on biodiversity

Oswaldo E. Sala

The greater the plant diversity in an ecosystem, the greater the ecosystem's productivity. A new analysis indicates that the higher productivity results from complementary patterns of species resource use.

Human activities are drastically altering Earth's biodiversity. To get a handle on what the consequences might be, ecologists have been busily carrying out experiments. But interpreting such experiments has been confounded by the possible opera-

tion of two different causal mechanisms, with contrasting implications. This matter is tackled by Loreau and Hector on page 72 of this issue¹. They have devised a way of teasing apart the two effects, drawing upon a formulation — the Price equation — used

in evolutionary genetics, and have tested their approach on a large body of data from European grasslands.

The continuing changes in biodiversity are of great concern. Some 5–20% of species in many groups of organisms² have become extinct through human action, either direct or indirect. And given continuing changes — in land use, climate, nitrogen deposition and concentration of atmospheric CO₂, and increasing species invasions to the detriment of the existing inhabitants — such losses look likely to increase³. Hence arises the pressing question of how changes in biodiversity affect the functioning of ecosystems and their ability to provide goods and services for humans. How will primary production and nutrient cycling be affected? And will the capacity of ecosystems to sequester carbon and provide food, fibre and clean water be impaired?

Several large, controlled experiments have shown that primary production seems to be higher with greater biodiversity^{4–6}. But ecologists have struggled to distinguish between two alternative hypotheses to account for the results. These are 'species complementarity' and the 'sampling effect'^{7–9}. The first is an ecological phenomenon, the second a statistical consequence of experimental design. What Loreau and Hector¹ have done is to propose a way of telling the difference between the two.

The species-complementarity hypothesis is based on the idea of a trade-off between species traits¹⁰. For example, different species may have deep or shallow roots, growth optima at high or low temperatures, or high or low relative growth rates with corresponding resistances to stress. Ecosystems that have a larger number of species will probably have a broader range of traits — and thus, for example, are more likely to draw on both shallow and deep layers of the soil, or to fix carbon at both low and high temperatures.

The sampling effect, in contrast, states that increased productivity with increasing diversity is an artefact of experimental design. It stems from the principle that different species are differentially adapted to a given environment. In most biodiversity experiments, the probability of a given sample containing the best-adapted species increases as diversity increases, because the species composition of each sample is a random draw from a finite pool of species.

Loreau and Hector's method¹ is intended to distinguish between species complementarity and the sampling effect. It is based on the Price equation, which is used in evolutionary genetics to calculate changes in the frequency of a trait between individuals in one generation and the ancestral generation as a function of the covariance between fitness and trait value¹¹. Loreau and Hector's insight is to have identified parallels between

the sampling effect in biodiversity experiments and the selection effect seen in the classic Price equation. Their new 'biodiversity equation' includes two terms that partition the 'biodiversity effect' into the distinct species-complementarity and sampling mechanisms. The biodiversity effect is the increase in yield with increasing biodiversity, yield in this case representing biomass, primary production or any other measure of ecosystem functioning.

The sampling effect is calculated as the covariance between a species' yield in monoculture and its yield in mixed plots. The species-complementarity effect is calculated as a function of the increase in yield of a mixture of species relative to the expected yield based on the yield of the same species growing in monocultures.

A hypothetical example is depicted in Fig. 1. This shows how yields of mixtures of plants can be higher than the sum of their yields in monoculture as a result of either species complementarity (Fig. 1a, b) or the sampling effect (Fig. 1c, d), and also how the contribution of each mechanism can be quantified using the Loreau–Hector equation. An explanation that invokes species complementarity for the increase in yield

might be that, in a water-limited ecosystem, two species use different water sources (say, shallow or deep soil layers). A sampling-effect explanation might be that species A is better adapted than species B, not only achieving a higher yield in monoculture but also dominating the mixed-species plots to the virtual exclusion of B.

Loreau and Hector¹ applied their equation to results from BIODEPTH⁶. This experiment, designed to assess the relationship between biodiversity and ecosystem functioning, had the same design (4 or 5 levels of plant-species diversity with 1 to 32 species) replicated in seven European countries. Loreau and Hector found that the main mechanism behind the increase in production with increasing biodiversity was species complementarity. This suggests that species losses, such as those expected in the near future³, may result in lower production and less effective resource use. For example, reduced uptake of soil nitrogen may lead to higher concentrations of nitrate below the root zone; such an effect has been observed in North America⁴, and could result in other environmental problems. Similarly, losses of biodiversity may hamper the ability of ecosystems to sequester carbon: it has been

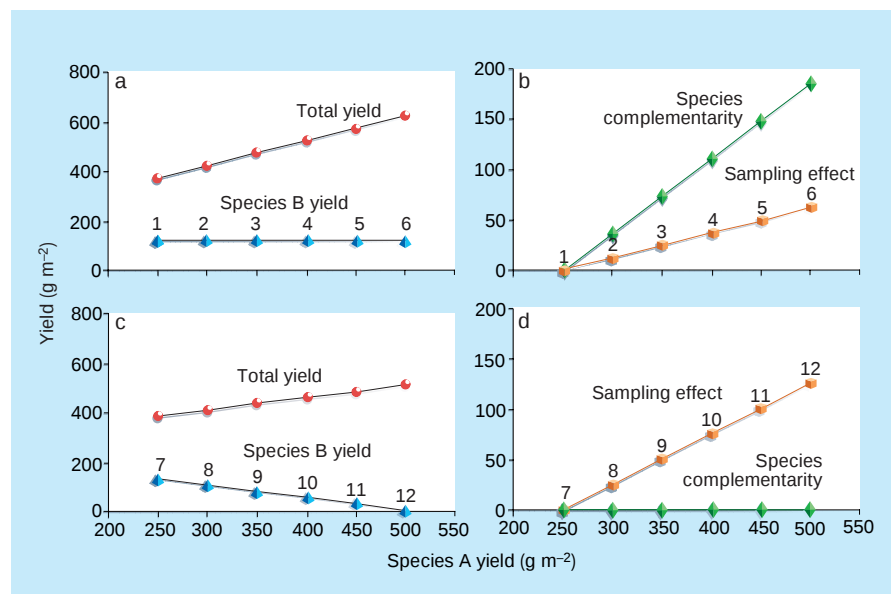


Figure 1 Examples of species complementarity and the sampling effect. These are the two possible explanations for the increased ecosystem yield that results from an increase in species diversity — in this case from one (monoculture) to two species (mixture). Shown here are various yields of 1:1 mixtures of species A and B growing together, compared with monoculture yields of 500 g m⁻² (species A) and 250 g m⁻² (species B). Each pair of data points (1–6, 7–12) represents different levels of yield of the mixtures. **a**, Species B maintains a constant yield of half of its monoculture yield (125 g m⁻²); yields of species A are 250–500 g m⁻². So, except in case 1 (each species has half of its monoculture yield), the mixture yield is greater than the sum of the monoculture yields. **b**, Application of the Loreau–Hector equation¹ for mixture yields 1–6 shown in **a**, to separate species complementarity and the sampling effect. Species complementarity predominates. **c**, Total yield in each instance of the mixture (except for case 7) is again larger than the sum of the monoculture yields. But here it is because of a shift in species dominance (A captures a large fraction of the resources). **d**, Application of the Loreau–Hector equation for mixture yields 7–12 shown in **c**. The sampling effect accounts entirely for the observed increase in yield of the mixture over that expected from the two monocultures.

shown¹² that plant species diversity controls the magnitude of the increase in carbon fixation as levels of atmospheric CO₂ increase.

How general is species complementarity? I suggest that the strength of this mechanism is related positively to the length of evolutionary history, and negatively to the frequency and intensity of disturbances to an ecosystem. Complementary resource use and synergistic relationships are more likely to occur among species that have had a chance to coevolve over long periods of time¹³. Frequent disturbance will prevent the evolution of tight differentiation in resource use, and will perturb or destroy symbiotic relationships. BIODEPTH was carried out using grassland species, mostly at sites where the potential natural vegetation was forest. These sites were maintained as grasslands because of frequent human intervention; if they had not been mowed once or twice a year, they would have reverted to forest. Species complementarity may act even more strongly in ecosystems that have been disturbed less often and have a longer evolutionary history.

We clearly need a better understanding of the relationships between biodiversity and

ecosystem functioning. There are two ways forward. The first is to apply this new tool, the Loreau–Hector equation, to other existing data sets, to see how general the species-complementarity principle is. The second is to gather — and then likewise analyse — fresh data for other ecosystems by carrying out experiments such as BIODEPTH in other areas of the world with different evolutionary and disturbance histories. ■

Oswaldo E. Sala is in the Department of Ecology and IFEVA, Faculty of Agronomy, University of Buenos Aires-CONICET, Avenida San Martín 4453, Buenos Aires C1417 DSQ, Argentina. e-mail: sala@ifeva.edu.ar

1. Loreau, M. & Hector, A. *Nature* **412**, 72–76 (2001).
2. Chapin, F. S. et al. *Nature* **405**, 234–242 (2000).
3. Sala, O. E. et al. *Science* **287**, 1770–1774 (2000).
4. Tilman, D., Wedin, D. & Knops, J. *Nature* **379**, 718–720 (1996).
5. Tilman, D. et al. *Science* **277**, 1300–1302 (1997).
6. Hector, A. et al. *Science* **286**, 1123–1127 (1999).
7. Tilman, D., Lehman, C. & Thomson, K. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
8. Huston, M. A. *Oecologia* **110**, 449–460 (1997).
9. Huston, M. A. et al. *Science* **289**, 1255 (2000).
10. Orians, G. H. & Solbrig, O. T. *Am. Nat.* **111**, 677–690 (1977).
11. Frank, S. A. *J. Theor. Biol.* **175**, 373–388 (1995).
12. Reich, P. et al. *Nature* **410**, 809–812 (2001).
13. MacArthur, R. *Geographical Ecology: Patterns in the Distribution of Species* (Princeton Univ. Press, Princeton, 1984).

Planetary science

Climate change on Venus

Ronald G. Prinn

Earth's climate has changed significantly over the past several million years. New theoretical work suggests that the climate of our nearest neighbour, Venus, may have also changed on similar timescales.

Venus is a most inhospitable planet. Its average surface temperature of 735 K is some 435 K higher than that of Earth. It has a thick atmosphere of carbon dioxide that exerts a surface pressure about 92 times greater than Earth's. Its craters and volcanoes are completely shrouded by thick clouds of sulphuric acid, and its surface features are revealed only in radar images. Not surprisingly, it has no oceans and no known life. But has this extreme climate always been the same, or does it change from millennium to millennium? In an article in *Icarus*, Mark Bullock and David Grinspoon¹ describe a numerical simulation of venusian climate that suggests it has oscillated over the past billion years between periods of global cooling and global warming.

Bullock and Grinspoon¹ have developed a new radiative-convective model of the venusian climate. It is based on recent data from spacecraft (particularly the 1990–1994 Magellan mission) and from ground-based telescopes, which together provide information on the geology, geophysics and atmospheric chemistry of Venus. Their model

is the first to use high-temperature, high-resolution spectroscopic data on the absorption properties of the major greenhouse gases found on Venus (mainly CO₂ with trace amounts of H₂O and SO₂). The authors also include data on the rates of reaction of these gases with surface minerals at high temperatures — reactions that limit their abundance in the atmosphere. They couple their climate model to models of cloud microphysics, volcanic outgassing of sulphur dioxide and water from the crust, surface chemistry, and water loss due to hydrogen atoms escaping from the high atmosphere into space.

The Bullock–Grinspoon¹ model indicates that between 600 million and 1,100 million years ago, Venus was cooler than it is today. It was cooler because sunlight was reflected by thick clouds of sulphuric acid (H₂SO₄) produced during a geologically active period when erupting lavas from global volcanic activity resulted in the build-up of SO₂ and H₂O in the atmosphere. This was followed by a period of warming as the SO₂ responsible for creating the clouds was depleted by reactions with minerals at the

surface, raising temperatures to 900 K. But the contribution of water vapour to greenhouse warming was subsequently lowered by the steady loss of hydrogen into space and the loss of oxygen through oxidation of surface minerals. This helped to cool Venus down to today's temperatures.

I first became interested in climate change on Venus in the early 1980s, spurred on by the intriguing results from the Pioneer mission to Venus in 1978. My own studies were aimed at understanding the processes that maintain sulphuric acid clouds on Venus, and the possibility that the clouds, and hence climate, could change as a result of changes in the emission of sulphur gases through volcanism and thermally driven surface chemistry^{2–4} (Fig. 1). Work in the laboratory indicated that SO₂, the precursor of sulphuric acid, could be removed from the atmosphere by reactions with surface minerals in 1.9 million years⁵ — a relatively short timescale for geological processes. And because the removal rate of SO₂ (and hence of H₂SO₄) increases with temperature, there is also the possibility of amplifying any warming or cooling trend.

The starting point for Bullock and Grinspoon's study was the Magellan mission to Venus in the 1990s. Magellan used radar to penetrate the clouds to produce, among other things, the first extremely high-resolution spatial map of the surface of Venus. This map indicated that the density of impact craters, and hence the number of comet and asteroid collisions recorded on the surface of Venus, was fairly low, suggesting that the present surface is only 600 million to 1,100 million years old⁶. The previous surface must have been obliterated by erupting magmas from volcanic activity on a global scale.

Bullock and Grinspoon's work indicates that H₂O and SO₂ have both cooperative and competitive effects on the venusian climate. The climate on Venus today is controlled by two main processes: global warming, largely resulting from the greenhouse effect of CO₂, and cooling, owing to the reflection of solar radiation by the thick clouds of sulphuric acid. Large increases in H₂O above today's levels could amplify the greenhouse warming effect and lead to thinning of the clouds through evaporation of their lowest layers. Overall, this could increase surface temperatures by 200 K. But large increases in SO₂ could cool the planet by up to 40 K by thickening these same clouds and increasing their reflectivity.

The authors propose that global volcanic activity 600 million to 1,100 million years ago injected large quantities of H₂O and SO₂ into the atmosphere. This thickened the clouds of sulphuric acid, and the resulting cooling was greater than any warming these gases contributed through

Daedalus

Learning to forget

Last week Daedalus contemplated the sensitivity of the brain to microwaves. He recalled brain theory, insofar as it exists. Each brain cell has many inputs (dendrites) and one main output (its axon); these are connected to other brain cells. Very plausibly, each dendrite is potentiated or inhibited by a specific protein molecule, whose configuration turns it on or off. And the shuffling of amino acids and the re-ordering of protein chains occurs at microwave frequencies. So mobile phones and microwave towers do a lot of unintended damage.

But controlled memory loss could be very welcome in psychiatry. Many people are haunted by dire memories, which prevent them re-entering places or circumstances, or make them fearful of certain normal human activities. A simple microwave irradiation, neatly erasing the damaging memory, could be gladly accepted. So DREADCO biochemists are irradiating test organisms, hoping to find that pattern of frequency or combination of frequencies that can erase a specific memory — in rats, for example, how to run a given maze.

The extension of this scheme to human psychiatry will be fraught with hazard. Fortunately, the main complaints about mobile phones refer to the loss of short-term memory only. Psychiatrists could screen mobile-phone users for the sort of information lost, and employ subjects who do not mind losing useless short-term data (perhaps certain recent breakfast menus). Furthermore, some quite worrying methods have been accepted by psychiatrists despite their potential for memory loss (electroconvulsive therapy is an example). So careful microwave-irradiation stands a good chance.

But Daedalus's main worry is the adoption of his methods by the politicians. A wide-band high-energy microwave assault on the brain might scramble all its data long before heating set in. The most convinced capitalist or advocate of human freedom could crumble before such brain-washing. Against that, how many fierce species have their hatred of humanity stored as software rather than hardware? A DREADCO selective irradiation might produce genuinely friendly lions and tigers, cheerfully accepting crocodiles and alligators. The range of tameable, tractable species could be wonderfully extended. Even the wild cat of the Scottish Highlands, said to hate everything on sight, might be converted to a purring fireside moggy.

David Jones

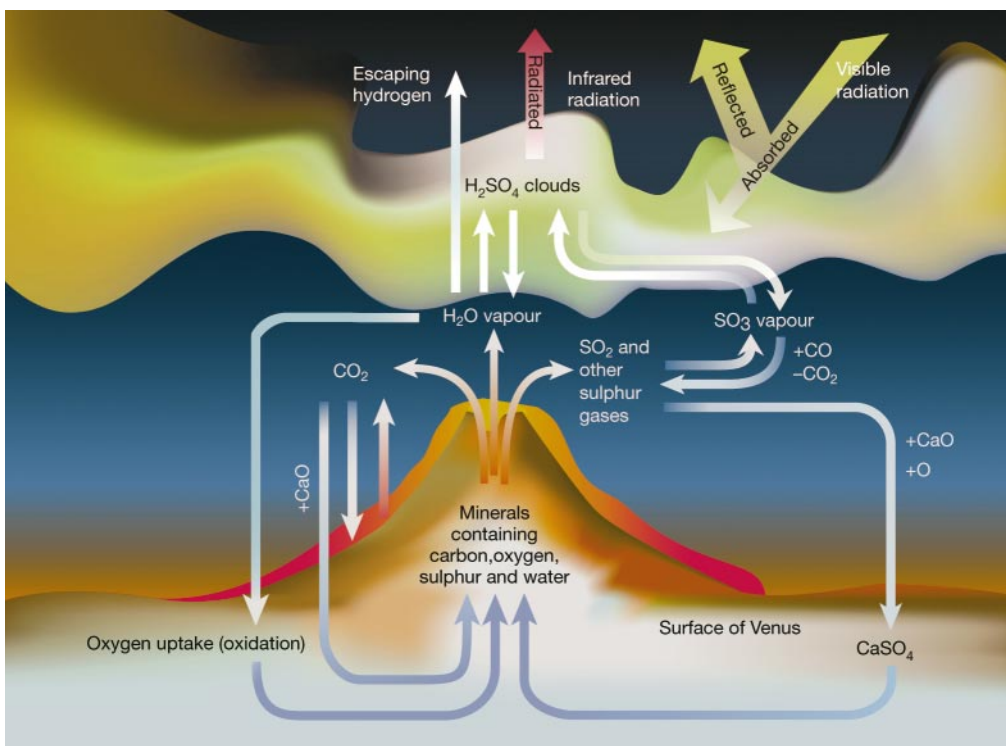


Figure 1 Key chemical and physical processes controlling the climate of Venus over long timescales^{1–5}. Venus has an atmosphere of 96.5% CO₂, which is primarily responsible for its greenhouse effect and high surface temperature. Venus also has a thick layer of sulphuric acid (H₂SO₄) clouds that reflect sunlight away from its surface, helping to cool it. The greenhouse warming is greater than the cooling effect of the clouds, making the surface of Venus much warmer than on Earth. A new numerical model developed by Bullock and Grinspoon¹ suggests that over the past 1 billion years the climate on Venus has experienced periods of both cooling and warming, largely triggered by global volcanic activity spewing out large amounts of sulphur dioxide (SO₂) and water vapour (H₂O).

the greenhouse effect. But after this period, the Bullock–Grinspoon model indicates relatively rapid removal of excess SO₂ by reaction with carbonates and other surface rocks, so the clouds became thinner and the planet warmed by 100 K. But the warming was limited by the steady loss of H₂O as hydrogen escaped into space and oxygen was lost at the surface, allowing Venus to cool again to its current temperature.

How stable is the climate of Venus now? Maintaining current levels of SO₂, and hence clouds of sulphuric acid, requires volcanic activity. Large impacts from comets could also bring water and sulphur to Venus, but it seems that in the near future the venusian climate will be a slave to the volcanism needed to sustain current levels of H₂O and SO₂ against their loss at the surface and into space. This simple picture is complicated by the possibility that changes in surface temperature can feed back to the processes involved in volcanism, by affecting the heat flow, and hence temperatures, deep below the surface⁷. So the story will be complex.

Such models are always speculative, but Bullock and Grinspoon suggest some specific space missions and observations that could clarify the existing picture. Measure-

ments of noble gases in the atmosphere could help quantify the flow of volatile gases from the interior into the atmosphere and space⁸. Close-up images of the surface — perhaps where the crust breaks in an impact — could reveal subsurface layers indicative of past climate change, like those seen on Mars⁹. And measurements of the flow in the atmosphere of visible and infrared radiation involved in the greenhouse effect, and the key gases involved in the chemical cycles⁴, would help enormously. Stay tuned for better forecasts of the climate on Venus.

Ronald G. Prinn is in the Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139-4307, USA.

e-mail: rprinn@mit.edu

1. Bullock, M. A. & Grinspoon, D. H. *Icarus* **150**, 19–37 (2001).
2. Prinn, R. G. *Scient. Am.* **252**, 46–54 (1985).
3. Von Zahn, U., Kumar, S., Niemann, H. & Prinn, R. in *Venus* (eds Hunten, D. H. et al.) 299–430 (Univ. Arizona Press, Tucson, 1983).
4. Prinn, R. G. in *Recent Advances in Planetary Meteorology* (ed. Hunt, G.) 1–15 (Cambridge Univ. Press, 1985).
5. Fegley, B. & Prinn, R. G. *Nature* **337**, 55–58 (1989).
6. Schaber, G. G. et al. *J. Geophys. Res.* **97**, 13257–13302 (1992).
7. Solomon, S. C., Bullock, M. A. & Grinspoon, D. H. *Science* **286**, 87–89 (1999).
8. Pepin, R. O. *Icarus* **92**, 2–79 (1991).
9. Treiman, A. H., Fuks, K. H. & Murchie, S. J. *Geophys. Res.* **100**, 26339–26344 (1995).

Obituary

Wang Ying-lai (1907–2001)

Wang Ying-lai, a biochemist recognized as the first scientist to engineer synthetic insulin, died in his sleep on 5 May 2001 in a hospital in Shanghai, China. He was 93 years old.

Wang was born in 1907 in the remote village of Shanhou on the island of Jinmen (Quemoy), off the coast of Fujian, China. Although he lost his father at the age of two, and his mother when he was six, he pursued his education vigorously against all odds during a period of wars and political turmoil in China in the 1920s and 1930s. After graduating from Jinling University (now the University of Nanjing) with a degree in chemistry, he went to the University of Cambridge in 1938. His mentor, David Keilin, quickly recognized his talent, and Wang earned his PhD in 1941. He was invited to stay at Cambridge to teach and to continue his research at the Dunn Nutritional Laboratory, moving to Cambridge's Molteno Institute in 1944.

When the Second World War ended, Wang decided to return to China, against the advice of both Keilin and Professor Joseph Needham, the foremost historian of Chinese science. Wang's aim was to help the country to develop a world-class research base in science, and his first job was a research professorship at the Medical College of the National Central University. In 1948, he became a senior member of the Medical Research Institute of what was then the Academia Sinica (now the Chinese Academy of Science) — the most prestigious research institution in China. Shortly after the liberation of China in 1949, he was appointed deputy director of the newly established Institute of Physiology and Biochemistry of the Academia Sinica of the People's Republic of China.

From 1958 until he retired in 1984, Wang founded and headed the Institute of Biochemistry of the Academia Sinica in Shanghai. It was in this capacity, according to an essay he wrote in the *People's Daily* in 1991, that he recruited several prominent Chinese scientists who were until then working in other countries. Over several years he helped to plan and develop the national agenda for biochemistry and molecular biology. He also set up several national training programmes to recruit and train hundreds of young scientists. He founded, and was for many years the editor-in-chief of, China's premier journal, the *Journal of Biochemistry and Biophysics*.



Pioneer in the total chemical synthesis of biological molecules

By far the most significant of Wang's scientific contributions was the synthesis of crystalline insulin. This hormone, produced in the pancreas, controls the sugar content in the blood and is widely used to treat diabetes. Under Wang's leadership, scientists in the 'collaboration team on the synthesis of insulin' achieved the synthesis in 1965. Although similar attempts were being made in the United States and Europe, Wang's group was ahead of the game.

Beginning the project in August 1958, the first task of Wang's team was to synthesize the 20 amino acids — the fundamental building blocks of any protein. This enormous job also involved the separation of the D- and L-stereoisomers of each amino acid (only the L-isomers are found in proteins). Then, using these building blocks, the team produced the so-called A and B amino-acid chains of insulin. Wang later said that the greatest challenge was to align the six cysteine amino acids so that they would form the correct disulphide bonds — one within the A chain, and two between the chains. Any mismatch would lead to an inactive product. Biological assays and X-ray crystallography confirmed the authenticity of the chemically synthesized insulin.

The total synthesis of insulin from chemically synthesized amino acids represented a conceptual breakthrough in converting lifeless chemicals to a protein with biological activity. This was accomplished before the development of solid-state peptide synthesis, for which

R. Bruce Merrifield was to receive the Nobel Prize in Chemistry in 1984. In the 1970s, a team headed by Wang achieved the total chemical synthesis of another biologically significant molecule — a transfer RNA (tRNA), which is involved in protein synthesis *in vivo*. This led to Wang's lifelong interest in tRNA synthetases, the enzymes that produce tRNAs.

Despite Wang's successful work with insulin, China — soon in the throes of the ten-year-long Cultural Revolution and other political upheavals — effectively denied him the opportunity of being nominated for the Nobel Prize in Chemistry. During that period, scientists in China were abruptly banned from conducting research and from contact with researchers in other countries. Swedish scientist A. Tiselius (a member of the Nobel Prize committee) and Chinese-American Nobel laureate C. N. Yang independently suggested that Wang be nominated for the prize at that time. About the decade of lost opportunity, Wang said in an interview with the *Straits Times* of Singapore in 1986, "We were like the proverbial hare which took a long nap while others were not like the tortoise." Nevertheless, he won international acclaim in 1966, before the Cultural Revolution began, when he presented his findings on insulin at a European biochemistry conference in Warsaw. And in 1988, ten years after Deng Xiaoping introduced new policies of economic reform and openness, Wang was allowed to receive a special achievement award at the Miami Biotechnology Symposium.

Wang was a patriotic, humble and dedicated scientist. In his 1991 essay he wrote: "Science is the first productive force. To transform China into a modern socialist power, we must strengthen science and technology... This is the responsibility of every Chinese." Asked by the *Straits Times* about the honours he had received, he replied, "The honours should go to everyone in the team, including 30 members in our institute and many from five other institutes." Despite the hardship he endured during the Cultural Revolution, he harboured no bitterness.

He is survived by his two sons, Jia-hu and Jia-nan; his daughters-in-law, Huang Jin-pei and Zhang Hong-xia; and two grandchildren, Wei-zhen and Wei-xian.

L. Ling-chi Wang

L. Ling-chi Wang is in the Department of Ethnic Studies, 506 Barrows Hall, University of California, Berkeley, California 94720-2570, USA.

e-mail: lcwang@uclink2.berkeley.edu

Surrogate mother for endangered *Cupressus*

A rare cypress tree increases its chances by using a clever reproductive strategy.

In higher plants, the embryo generally derives from the fusion of male and female gametes, although it may sometimes develop from only female cells. Here we show how the diploid pollen of the Mediterranean cypress tree *Cupressus dupreziana* naturally produces an embryo without fertilization that is nourished in the seed tissues of a surrogate mother, *Cupressus sempervirens*. This reproductive strategy of paternal apomixis, which to our knowledge has not been seen before in plants, could be an adaptation by this species in response to the threat of extinction.

The life cycle of plants generally alternates between two generations: the diploid sporophyte that produces haploid spores by meiosis, and the haploid gametophyte that produces gametes. Fertilization of a female gamete by a male gamete gives rise to a zygote that produces an embryo, the new sporophyte. However, in apomictic plants the embryo develops from maternal cells without fertilization¹. Apomixis has been frequently reported in angiosperms but was

thought not to occur in gymnosperms².

We have previously discovered significant anomalies in the reproductive structures of a Mediterranean gymnosperm, *Cupressus dupreziana* A. Camus (Fig. 1). In this variant, viable pollen (male gametophyte) is diploid³, embryos do not have the same allozymes as their mothers⁴, and the endosperm (seed nutritive tissue) is not haploid⁵, although in gymnosperms it is derived solely from the female gametophyte⁶. These anomalies suggested that the embryo results from the development of diploid pollen⁴ and led us to examine six 15-year-old families produced by controlled crosses of *Cupressus sempervirens* L. (as female) $\times$ *C. dupreziana* (as male). *C. sempervirens* is native to the eastern Mediterranean basin and has been widely propagated; *C. dupreziana* is native to the Tassili N'Ajjer desert of Algeria⁷ and is one of the most threatened trees in the world.

We compared the characteristics of *C. sempervirens* $\times$ *C. dupreziana* progeny to those of their parents by using the following morphological and cytological traits to differentiate the two species^{3,5,8}: orientation of terminal twigs (in one plane in *C. dupreziana*; in all directions in *C. sempervirens*); female cone size (larger in *C. sempervirens*); percentage of filled seeds (always low in *C. dupreziana*); endosperm ploidy levels (only even levels: $2n$, $4n$, $6n$... in *C. dupreziana*); pollen diameter ($38\text{ }\mu\text{m}$ in *C. dupreziana*; $28\text{ }\mu\text{m}$ in *C. sempervirens*); and pollen ploidy level (diploid in *C. dupreziana*). For all these traits, all progeny were identical to the male tree, *C. dupreziana*.

We assessed genetic diversity using two types of marker: isozymes (seven polymorphic systems: Fest, Idh, Lap, 6PgD, Pgi, Pgm and Skdh) in one *C. sempervirens* $\times$ *C. dupreziana* family, and random amplification of polymorphic DNA (RAPD; four operon primers: OPA-08, OPA-15, OPA-18 and OPR-07) in four families. A biparental, codominant inheritance was previously reported in *C. sempervirens* for these isozymes⁹, whereas the genetic control of the RAPD markers was unknown. The markers allowed identification of all the parents. Progeny had a single genetic pattern that was strictly identical to that of the father, *C. dupreziana* (Fig. 2).

Our results confirm at the interspecific level our hypothesis that pollen development in *C. dupreziana* is apomictic. This leads to the production of embryos that are genetically unrelated to the other seed components (maternal sporophyte and

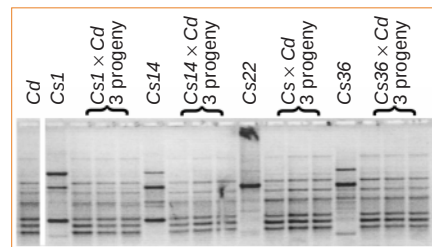


Figure 2 Genetic diversity, as revealed by random amplification of polymorphic DNA markers, among one *Cupressus dupreziana* tree used as male (Cd), four *Cupressus sempervirens* trees used as females (Cs1, Cs14, Cs22 and Cs36) and 12 progeny produced by the four *C. sempervirens* $\times$ *C. dupreziana* controlled crosses. The operon primer OPA-18 reveals different patterns for the five parents but produces a monomorphic profile for all progeny. This profile is similar to that of the male parent, indicating a strictly paternal origin of DNA.

gametophyte). We have also shown here that another cypress species can be a surrogate mother for this embryogenic pollen. This is, to our knowledge, the first report of paternal apomixis in plants. We suspect that this deviant reproductive pattern evolved in response to the reduction of *C. dupreziana* population size, which is today limited to 231 individuals. Inbreeding in small populations of naturally outbreeding species reduces fitness and increases the risk of extinction¹⁰. Apomixis, as well as other factors that limit inbreeding, may therefore confer a selective advantage.

Christian Pichot*, Mohamed El Maâtaoui†, Sabrina Raddi‡, Paolo Raddi§

*INRA, Unité de Recherches Forestières Méditerranéennes, Avenue Vivaldi, 84000 Avignon, France
e-mail: pichot@avignon.inra.fr

†UMR INRA/UAPV Ecologie des Invertébrés, Département de Biologie, Faculté des Sciences, 33 Rue L. Pasteur, 84000 Avignon, France

‡DISTAF, Florence University, Via S. Bonaventura 13, 50145 Florence, Italy

§Institute of Forest Tree Pathology, CNR, Piazzale Delle Cascine 28, 50144 Florence, Italy

- den Nijs, A. P. M. & van Dijk, G. E. in *Plant Breeding: Principles and Prospects* (eds Hayward, M. D., Bosemark, N. O. & Romagosa, I.) 229–245 (Chapman & Hall, London, 1993).
- Mogie, M. *The Evolution of Asexual Reproduction in Plants* (Chapman & Hall, London, 1992).
- Pichot, C. & El Maâtaoui, M. *Theor. Appl. Genet.* **101**, 574–579 (2000).
- Pichot, C., Fady, B. & Hochu, I. *Ann. For. Sci.* **57**, 17–22 (2000).
- Pichot, C., Borrut, A. & El Maâtaoui, M. *Sex. Plant Reprod.* **11**, 148–152 (1998).
- Singh, H. *Embryology of Gymnosperms* (Gebrüder-Borntraeger, Berlin, 1978).
- Barry, J. P. et al. *Bull. Soc. Hist. Nat. Afr. Nord* **61**, 95–178 (1970).
- Gaussen, H. *Les Gymnospermes Actuelles et Fossiles* (Trav. Lab. Forest., Toulouse, 1968).
- Papageorgiou, A. C., Bergmann, F., Gillet, E. & Hattemer, H. H. *Silvae Genet.* **42**, 109–111 (1993).
- Frankham, R. & Ralls, C. *Nature* **21**, 441–442 (1998).



Figure 1 The cypress *Cupressus dupreziana* is native to the inhospitable Tassili N'Ajjer desert of Algeria and is one of the world's most endangered trees, with only 231 remaining.

Pathogen reservoirs

Chironomid egg masses and *Vibrio cholerae*

Cholera is a severe diarrhoeal disease triggered by a toxin produced by specific biotypes of the bacterium *Vibrio cholerae* that are pathogenic only to humans. There seems to be no chronic state of the disease and the natural reservoir of the pathogen is environmental¹. Here we show that egg masses of the non-biting midge *Chironomus* sp. (Diptera) harbour *V. cholerae* and act as its sole carbon source, thereby providing a possible natural reservoir for the cholera bacterium.

Chironomids are the most widely distributed and often the most abundant insect in fresh water². Females deposit egg masses, each containing hundreds of eggs encased in a layer of gelatin, at the water's edge — a convenient location for bacteria to exploit this nutrient-rich substrate.

Two hundred floating *Chironomus* egg masses (Fig. 1a, b) collected from a waste-stabilization pond settled out overnight as thousands of individual eggs (Fig. 1c), most of which did not hatch. To test whether bacteria feeding on and destroying the gelatin matrix might account for this phenomenon, we extracted freshly collected egg masses with a minimal salt-solution medium containing no carbon source³ in order to isolate and identify any bacteria present.

Four isolates were selected and identified as *Vibrio cholerae* non-O1 non-O139 by standard microbiological tests, as well as by their serotype and fatty-acid profiles. We then collected *Chironomus* egg masses from other waste-stabilization ponds in various

regions of Israel and were able to isolate *V. cholerae* from all of these samples.

We inoculated fresh egg masses in a salt-solution medium with 1×10^3 *V. cholerae* per ml. Two controls were run concurrently: egg masses incubated alone under the same conditions, and medium without egg masses inoculated with 1×10^3 per ml *V. cholerae*. We determined growth of *V. cholerae* after 24 h by plating samples on thiosulphate–citrate–bile salt substrate.

The number of bacterial colony-forming units (CFUs) that developed in the first control was always less than 0.1% of that recovered in the treated samples; CFUs recovered in the second control did not change. In the medium containing egg masses as the sole carbon source, *V. cholerae* reached 2×10^6 CFU ml⁻¹. Similar results were obtained when sterilized egg masses were provided as a carbon source. These findings show that egg masses can provide a carbon source to support the development and multiplication of *V. cholerae*.

Although the *V. cholerae* biotypes isolated here are non-pathogenic, it is likely that chironomid egg masses would also be a suitable (and abundant) substrate for the pathogenic *V. cholerae* O1 and O139, assuming that the microhabitat of the pathogenic biotypes is similar⁴.

An association has been noted between both viable and 'viable but non-culturable' *V. cholerae* and zooplankton, and copepods have been implicated in the spread of cholera^{5,6}. Propagules may be carried by marine zooplankton along the continental seashore, aided by climatic events such as the El Niño Southern Oscillation^{7–9}. These results are relevant to the dispersion of pandemics and to the autochthonous existence

of *V. cholerae* in endemic locales during periods between epidemics — when there is local build-up of the bacterium but no outbreak of disease. Our findings indicate that chironomid egg masses may serve as an intermediate 'host' reservoir for *V. cholerae*, facilitating its survival and multiplication in freshwater bodies.

Meir Broza, Malka Halpern

Department of Biology, Faculty of Sciences and Science Education, University of Haifa, Oranim, Tivon 36006, Israel

e-mail: broza@research.haifa.ac.il

1. Heidelberg, F. J. *et al.* *Nature* **406**, 477–484 (2000).
2. Armitage, P., Cranston, P. S. & Pinder, L. C. V. (eds) *The Chironomidae: the Biology and Ecology of Non-Biting Midges* (Chapman & Hall, London, 1995).
3. Echols, H., Garen, A., Garen, S. & Torriani, A. J. *Mol. Biol.* **3**, 425–438 (1961).
4. Lewin, S. M. in *Cholera and the Ecology of Vibrio cholerae* (eds Drasar, B. S. & Forrest, B. D.) 228–254 (Chapman & Hall, London, 1996).
5. Colwell, R. R. & Huq, A. in *Vibrio cholerae and Cholera: Molecular to Global Perspective* (eds Wachsmuth, I. K., Blake, P. A. & Olsvik, Ø.) 117–133 (ASM, Washington DC, 1994).
6. Roszak, D. B. & Colwell, R. R. *Microbiol. Rev.* **51**, 365–379 (1987).
7. Colwell, R. R. *Science* **274**, 2025–2031 (1996).
8. Colwell, R. R. & Huq, A. J. *Appl. Microsc. Symp.* **85** (suppl.), 134–137 (1999).
9. Harvell, C. D. *et al.* *Science* **285**, 1505–1510 (1999).
10. Nolte, U. *Ent. Scand. Suppl.* **43**, 5–75 (1993).
11. Broza, M., Halpern, M. & Inbar, M. *Wat. Sci. Technol.* **42**, 71–74 (2000).

Carbon fixation

Photosynthesis in a marine diatom

The first stable product of photosynthetic carbon fixation by land plants is either the three-carbon molecule phosphoglycerate (in C_3 plants) or the four-carbon compounds malate or aspartate (in C_4 and CAM (crassulacean-acid metabolism) plants). Reinfelder *et al.* infer that a C_4 biochemical pathway of carbon fixation also operates in marine diatoms^{1,2} on the basis of their discovery of the enzyme phosphoenolpyruvate (PEP) carboxylase and of their ¹⁴C-tracer results in the marine diatom *Thalassiosira weissflogii*. However, we consider that further analysis is called for to demonstrate that this marine diatom meets all the criteria for C_4 photosynthesis.

C_4 photosynthesis has several features that are essential to CO_2 -concentrating mechanisms in general³, such as an active, photosynthetically driven, CO_2 -capture system (PEP carboxylase), an intermediate pool of captured CO_2 (C_4 acids, such as malate), a mechanism to release CO_2 from this pool (decarboxylation of C_4 acids) and a compartment in which to concentrate CO_2 around ribulose 1,5-bisphosphate carboxylase–oxygenase (Rubisco) and to reduce its leakage (for example, the bundle sheath or a functionally equivalent structure). We question whether all these fea-

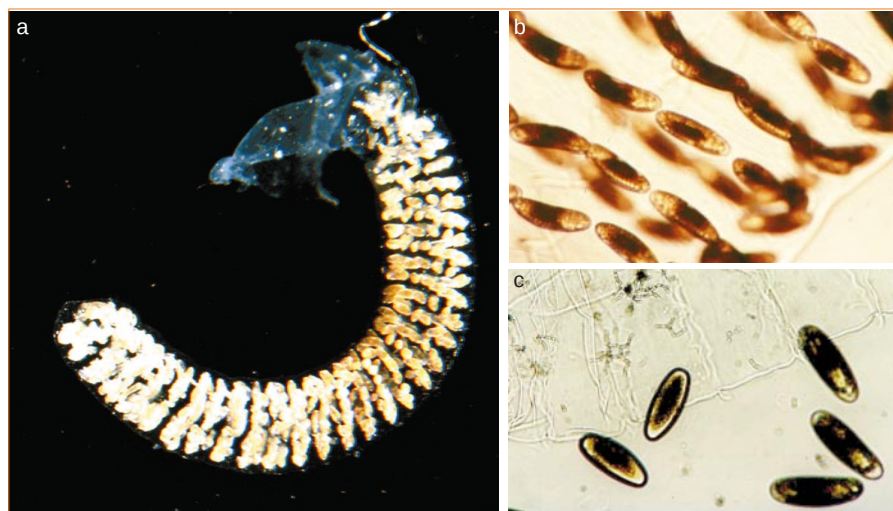


Figure 1 Egg mass of *Chironomus* before and after inoculation with *Vibrio cholerae*. The eggs are arranged in a row, folded into loops to form a spiral, and embedded in a thick, gelatinous cylinder. Egg masses of *Chironomus* spp. are found in freshwater habitats, where they can reach 25 × 5 mm in size and contain about 1,000 eggs. The appearance of several thousand egg masses at one site is not unusual and, in extreme cases, gelatinous layers several centimetres thick are visible from a distance^{10,11}. **a**, String-shaped mass of *Chironomus luridus* eggs (original magnification, × 4). **b**, Enlarged portion of the mass, showing the gelatinous matrix. **c**, Enlarged portion 5–10 h after exposure to *V. cholerae*. Eggs are no longer arranged in spiral rows but protrude from the gelatin, which has been partially consumed by the bacteria. Original magnification in **b** and **c**, × 40.

tures have so far been adequately demonstrated in *T. weissflogii*.

Considering first the nature and location of the carboxylase and decarboxylase, Reinfelder *et al.* suggest¹ that the significance of PEP carboxylase in diatoms may have been missed because of lack of carbon limitation during cell culture. In many studies, however, the dissolved CO₂ concentration was probably the least well controlled growth condition⁴, particularly for enzyme assays in which the requirement for algal biomass outweighed other considerations. Earlier reports of PEP carboxylase activity in diatoms were probably due in fact to PEP carboxykinase activity^{5,6}. The PEP carboxylase activity found by Reinfelder *et al.* in the chloroplast (53% of total) in the Percoll-purified chloroplast fraction of *T. weissflogii*¹, and the finding of 22% Rubisco activity in the soluble fraction, suggest considerable cross-contamination of the preparations, which would cast doubt on the localization of PEP carboxykinase.

Reinfelder *et al.* also base their conclusions on ¹⁴C-tracer kinetics. The shortest labelling time they observed using ¹⁴C is 5 s, with much higher labelling of malate than has previously been found in diatoms⁵. However, even the 65% labelling by ¹⁴C they report in the C₄ (β-carboxyl) position of malate is not sufficient to justify designation as C₄ biochemistry because secondary carboxylation of labelled PEP, arising from labelled phosphoglycerate generated by Rubisco, can occur.

One crucial diagnostic test for C₄ biochemistry is whether the label on malate extrapolates back to 100% at time zero. However, the two points provided in the data of Reinfelder *et al.*¹ are insufficient to establish a time course. The brown seaweed *Ascophyllum nodosum* (Phaeophyceae) is closely related to diatoms: it has a similar C₄-like photosynthetic physiology⁷ and shows significant ¹⁴C-labelling of C₄ compounds⁸. But under conditions of inorganic-carbon saturation (sea water), the initial ¹⁴C-incorporation product after 1–2 s of photosynthesis is phosphoglycerate⁷. Also, C₄ plants that use PEP carboxykinase as their decarboxylating enzyme use aspartate⁹ rather than malate¹ as a substrate.

The finding that phosphoglycerate is the first product of carbon fixation in photosynthesis by *Ascophyllum*⁷ and by various diatoms⁶ suggests that heterokont algae photosynthesize using a C₃ pathway and that synthesis of C₄ compounds using β-carboxylases (mainly, if not exclusively, PEP carboxykinase) has a ubiquitous anaplerotic role in these organisms. The increased labelling of C₄ compounds relative to phosphoglycerate under conditions of low inorganic carbon is also seen when growth rate is limited by (non-carbon) nutrients or by light⁶.

Invoking a role for C₄ photosynthesis to explain the physiological properties and ¹³C-discrimination values of diatoms seems unnecessary, given the capacity of such cells to accumulate inorganic carbon and hence CO₂ by means of the biophysical CO₂-concentrating mechanism — a process for which there is a large amount of evidence^{6,7,10}. Until these issues are unequivocally resolved and the other features of C₄ photosynthesis are demonstrated, we believe that it is premature to designate marine diatoms as C₄ photosynthesizers in the traditional sense.

Andrew M. Johnston*, John A. Raven†, John Beardall‡, Richard C. Leegood§

*Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

e-mail: ajohns@scri.scri.ac.uk

†Biological Sciences Institute, University of Dundee, Millers Wynd, Dundee DD1 4HN, UK

‡Department of Biological Sciences, Monash University, Clayton, Victoria 3800, Australia

§Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

1. Reinfelder, J. R., Kraepiel, A. M. L. & Morel, F. M. M. *Nature* **407**, 996–999 (2000).
2. Riebesell, U. *Nature* **407**, 959–960 (2000).
3. Badger, M. R. & Spalding, M. H. in *Photosynthesis: Physiology and Metabolism* (eds Leegood, R. C., Sharkey, T. D. & von Caemmerer, S.) 369 (Kluwer Academic, Dordrecht, 2000).
4. Johnston, A. M. & Raven, J. A. *Marine Ecol. Prog. Ser.* **87**, 517–521 (1992).
5. Beardall, J., Mukerji, D., Glover, H. E. & Morris, I. J. *Phycol.* **12**, 409–417 (1976).
6. Beardall, J. *Aquat. Bot.* **34**, 105–130 (1989).
7. Johnston, A. M. *Can. J. Bot.* **69**, 1123–1132 (1991).
8. Johnston, A. M. & Raven, J. A. *Phycologia* **26**, 159–166 (1987).
9. Sage, R. F. & Monson, R. K. (eds) *C₄ Plant Biology* 596 (Academic, San Diego, 1999).
10. Raven, J. A. *Adv. Bot. Res.* **27**, 85–209 (1997).

Reinfelder et al. reply — Johnston *et al.* are not convinced that a C₄ photosynthetic pathway exists in microalgae. The essential feature of C₄ photosynthesis is the fixation of inorganic carbon as a C₄ compound that is subsequently decarboxylated to provide CO₂ as a substrate for Rubisco in the Calvin cycle. It is known that many microalgae incorporate at least some inorganic carbon directly into C₄ compounds such as malate, but the controversy hinges on whether this carbon is released and fixed by Rubisco.

The extrapolation of ¹⁴C incorporated into the C₄ position of malate to 100% at time zero, which is proposed as a crucial diagnostic test by Johnston *et al.*, is thus not a direct verification of C₄ photosynthesis. It would be experimentally difficult to determine and is not mathematically well defined — for example, this approach yields an intercept of only 89% in maize, a well-known C₄ plant¹. A better demonstration is provided by the transfer of ¹⁴C from the C₄ compound malate to the C₃ compound phosphoglycerate, the first product of Rubisco, as we have shown. Also, we have recently found (unpublished results) that addition of the C₄ compound oxaloacetate

to compensated cultures of *T. weissflogii* triggers immediate production of O₂ in the light, but no O₂ consumption in the dark, which further supports our proposal that this diatom uses C₄ photosynthesis.

Johnston *et al.* also comment on the substrate of PEP carboxykinase and the localization of enzymes. However, oxaloacetate, which is the substrate for PEP carboxykinase, can be produced by oxidation of aspartate or malate. We agree that cellular-fractionation techniques cannot definitively establish how the enzymes of a photosynthetic C₄ pathway are compartmentalized — so far, fractionation results have been confirmed by other techniques in the case of carbonic anhydrase, which has been shown to be located in the cytoplasm of *T. weissflogii* using antibodies.

There is strong evidence that marine diatoms can concentrate inorganic carbon for photosynthesis^{2,3}, but little to indicate how such a mechanism might work. C₄ pathways are used to accomplish this in multicellular C₄ plants and so may serve the same function in marine diatoms.

John R. Reinfelder

Department of Environmental Sciences, Rutgers University, 14 College Farm Road, New Brunswick, New Jersey 08901-8551, USA

e-mail: reinfelder@envsci.rutgers.edu

1. Hatch, M. D. & Osmond, C. B. in *Encyclopedia of Plant Physiology* Vol. 3 (eds Stocking, C. R. & Heber, U.) 144–184 (Springer, Berlin, 1976).
2. Raven, J. A. *Adv. Bot. Res.* **27**, 85–209 (1997).
3. Tortell, P. D., Rau, G. H. & Morel, F. M. M. *Limnol. Oceanogr.* **45**, 1485–1500 (2000).

Aerodynamics

Insects can halve wind-turbine power

For no apparent reason, the power of wind turbines operating in high winds may drop, causing production losses of up to 25 per cent¹. Here we use a new flow-visualization technique to analyse airflow separation over the blades and find that insects caught on the leading edges in earlier low-wind periods are to blame. These potentially catastrophic power glitches can be prevented simply by cleaning the blades.

Unpredictable changes in power levels have been noted on wind farms in California, with power sometimes falling to half the output predicted from the turbine design and generating two or more different power levels at the same wind speed (Fig. 1a). Although this phenomenon (termed 'double' or 'multiple' stalling) has been investigated^{2–4}, the cause has remained unknown.

One study⁵ commissioned by a turbine manufacturer (NEG Micon) used a new invention called a stall flag⁶ (patent, Energy Centre of The Netherlands) as a flow-

separation detector (Fig. 1b) to try and solve the problem. This device works on the principle of a hinged flap that opens up in a separated airflow to uncover an individual reflector (Fig. 1c) which allows the flow to be visualized. Operation of a stall flag on a turbine with a rotor diameter of 44 metres is illustrated in Fig. 1d, in which the light tracks are from exposed reflectors and indicate where the blades stall.

We found that the stalling behaviour of the blades depends on the degree of contamination of the leading edges. However, the reduction in power should be continuous (as debris on the blades would be expected to accumulate gradually) rather than stepped in distinct levels as shown in Fig. 1a.

We considered the possibility that flying insects caught on the turbine blades could explain this effect. Insects prefer to fly in conditions of high air humidity, low wind and temperatures above about 10 °C. Under these circumstances, they will increasingly foul the leading edges of the blades. In low winds, the incident angle between the flow and the blades is small, which corresponds to low air velocity around the leading edges, so the blade is not susceptible to contamination of the leading edge and the power output is unaffected. Insects rarely fly in high winds, so turbines operating in steady high-wind conditions do not become contaminated and power levels remain constant.

In high winds, however, the angle between the flow and the blades increases and the aerodynamic suction peak (the area of minimum pressure and maximum air velocity) shifts to the leading edge. If this happens to be already spattered with dead insects, power output will fall: the greater the contamination at the suction peak, the sooner the blades will stall and the more lift will be lost (Fig. 1e). Thus after each period of low wind, the amount of insect contamination may change, causing a different power level to be produced in high wind.

We verified this hypothesis experimentally by using stall flagging to compare air-flow over smooth blades with that over blades that had been artificially roughened on their leading edges (by installing a zigzag tape of maximum thickness 1.15 mm). The two turbines used were within 50 metres of each other to ensure equal inflow (Fig. 1f). A 25-Hz digital video camera recorded the stall-flag signals, providing thousands of computer-processed images which indicated that flow separation on the roughened blades was significantly increased at wind speeds of 11–25 m s⁻¹. This effect extended over the entire blade span, which explains the previously observed power losses (Fig. 1a). Moreover, power output from rough and smooth-bladed turbines was equal at low wind speeds, but higher from the 'clean' blades at high wind speeds (Fig. 1g), neatly reproducing the effect shown in Fig. 1a.

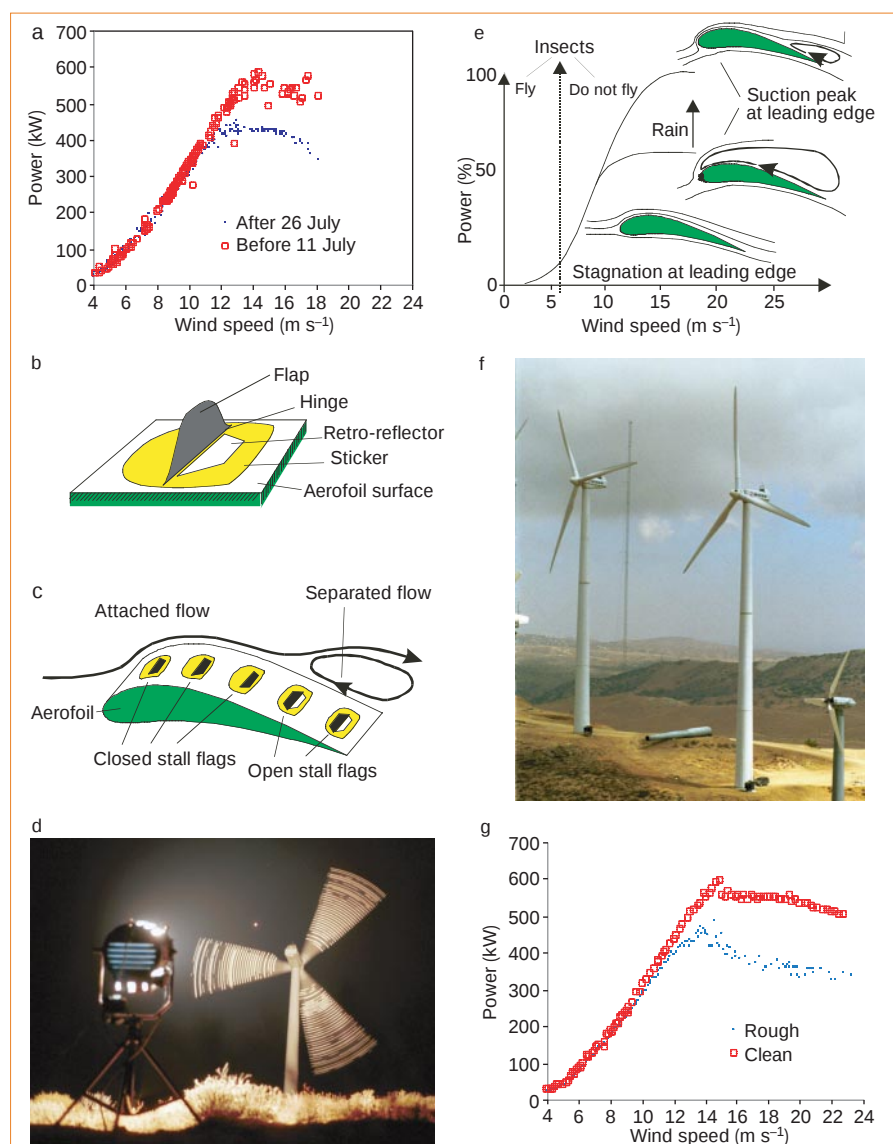


Figure 1 Insects cause multiple levels of power output from wind turbines. **a**, Example of production of two or more power levels at the same wind speed on different dates. **b**, The stall flag, consisting of a hinged flap and a reflector. **c**, Stall flags, showing the separated-flow area on an aerofoil. **d**, Recording of stall-flag signals from the NEG Micon turbine in California. The light tracks are produced by reflected light from open stall flags. **e**, Illustration of the insect hypothesis proposed to explain multiple power levels. **f**, The two turbines used for validation of the insect hypothesis; these were only 50 m apart to ensure equal air inflow. **g**, Power curves for the two turbines with 'rough' or 'clean' blades, which are similar to those in **a**.

We also studied a time series for the power output from four different turbines and found that the power at high wind speeds decreased markedly after every period of low wind speed, rising again after the blades were cleaned either manually or by rain, as expected. It is likely that accumulation of ice or dirt on the blades could create distinct power levels in high winds in the same way as insect contamination.

Gustave P. Corten*, Herman F. Veldkamp†

*Unit of Wind Energy, Energy Centre of The Netherlands, PO Box 1, 1755 ZG Petten, The Netherlands
e-mail: corten@ecn.nl

†NEG Micon A/S, Alsvej 21, 8900 Randers, Denmark

1. Dyrmoose, S. Z. & Hansen, P. *The Double Stall Phenomenon and How to Avoid It* (IEA Joint Action Aerodynamics of Wind Turbines, Lyngby, 1998).
2. Snel, H. et al. *Progress in the Joule Project: Multiple Stall Levels* (EUWEC Proc., Nice, 1999).
3. Bak, C. et al. *Wind Energ.* **2**, 195–210 (1999).
4. Corten, G. P. *Insects Cause Double Stall* (13th IEA Expert Meeting on the Aerodynamics of Wind Turbines, Stockholm, 1999).
5. Corten, G. P. *Flow Separation on Wind Turbine Blades*. Thesis, Univ. Utrecht (2001).

correction

Thin solid films roll up into nanotubes

O. G. Schmidt, K. Eberl
Nature **410**, 168 (2001)

Citation of additional earlier work by Prinz et al. (for example, see ref. 1) relating to these results was inadvertently omitted.

1. Prinz, V. Ya. et al. *Inst. Phys. Conf. Ser.* **166** Ch. 4, 199–202; 203–206 (2000).

1. Madsen, H. A. *Aerodynamics of a Horizontal Axis Wind Turbine in Natural Conditions* (Risø-M, Denmark, 1991).

Dopamine responses comply with basic assumptions of formal learning theory

Pascale Waelti*, Anthony Dickinson† & Wolfram Schultz*

* Institute of Physiology and Programme in Neuroscience, University of Fribourg, CH-1700 Fribourg, Switzerland

† Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK

According to contemporary learning theories, the discrepancy, or error, between the actual and predicted reward determines whether learning occurs when a stimulus is paired with a reward. The role of prediction errors is directly demonstrated by the observation that learning is blocked when the stimulus is paired with a fully predicted reward. By using this blocking procedure, we show that the responses of dopamine neurons to conditioned stimuli was governed differentially by the occurrence of reward prediction errors rather than stimulus–reward associations alone, as was the learning of behavioural reactions. Both behavioural and neuronal learning occurred predominantly when dopamine neurons registered a reward prediction error at the time of the reward. Our data indicate that the use of analytical tests derived from formal behavioural learning theory provides a powerful approach for studying the role of single neurons in learning.

Classic theories assume that predictive learning occurs whenever a stimulus is paired with a reward or punishment^{1,2}. However, more recent analyses of associative learning argue that simple temporal contiguity between a stimulus and a reinforcer is not sufficient for learning and that a discrepancy between the reinforcer that is predicted by a stimulus and the actual reinforcer is also required^{3–6}. This discrepancy can be characterized as a ‘prediction error’⁷. Presentations of surprising or unpredicted reinforcers generate positive prediction errors, and thereby support learning, whereas omissions of predicted reinforcers generate negative prediction errors and lead to reduction or extinction of learned behaviour. Expected reinforcers do not generate prediction errors and therefore fail to support further learning even when the stimulus is consistently paired with the reinforcer. Modelling studies have shown that neuronal messages encoding prediction errors can act as explicit teaching signals for modifying the synaptic connections that underlie associative learning^{8–17}.

Current research suggests that one of the principal reward systems of the brain involves dopamine neurons^{18–21}. Both psychopharmacological manipulations and lesions of the dopamine system impair reward-driven behaviour of animals^{18,21}, and drugs of abuse, which provide strong artificial rewards, act via dopamine neurons^{19,20}. Neurobiological investigations of associative learning have shown that dopamine neurons respond phasically to rewards in a manner compatible with the coding of prediction errors^{22–28}, whereas slower dopamine changes are involved in a larger spectrum of motivating events^{18,28,29}. Dopamine neurons show short-latency, phasic activations when rewards occur unpredictably, they are not modulated by fully predicted rewards and show phasically reduced activity when predicted rewards are omitted. During initial learning, when rewards occur unpredictably, dopamine neurons are activated by rewards. They gradually lose the response as the reward becomes increasingly predicted^{24,25,27}. The conditioned, reward-predicting stimulus starts to drive the dopamine neurons as they lose their response to the reward itself^{24,25}. Dopamine neurons also respond to novel, attention-generating and motivating stimuli²⁴, indicating that attentional mechanisms also contribute^{28,30}. Results from neuronal modelling suggest that the responses of dopamine neurons to primary rewards and conditioned stimuli may constitute particularly effective teaching signals for associative learning^{4,16,17,31,32} and embody the properties of the teaching signal of the temporal difference reinforcement model^{33–36}.

The present experiment explored how dopamine neurons of primates acquire responses to reward-predicting stimuli. In order

to determine the concordance between dopamine responses and reward prediction errors, we employed the canonical paradigm for assessing the role of prediction errors in learning, the blocking test³⁷. We demonstrated behavioural sensitivity to prediction errors, investigated the acquisition of neuronal responses to stimuli, and related behavioural and neuronal learning to the responses of dopamine neurons at the time of the reward.

Blocking and prediction errors

The blocking paradigm generated differential prediction errors by training a target stimulus in compound with a pretrained stimulus. As the pretrained stimulus has already been established as a predictor of the reinforcer, the reinforcer on compound trials is expected and therefore generates a minimal prediction error. Consequently, learning about the target stimulus is blocked. The blocking effect demonstrates that prediction errors, rather than simple stimulus–reinforcer pairings, play a crucial role in human conditioning³⁸, causal learning³⁹, animal conditioning³⁷ and artificial network learning³¹.

We employed a visual stimulus A which predicted the delivery of juice to the monkey (A+ trials), whereas a control stimulus B was not followed by reward (B– trials) (Fig. 1a). Learning was assessed by licking at a spout in anticipation of the reward in A+ but not B– trials. Reward delivery was not dependent on the animal’s anticipatory licks in this classical (Pavlovian) conditioning procedure. After training, the reward following stimulus A and the absence of reward following B were predicted and should not have generated prediction errors. During subsequent compound training, two stimuli (X and Y) were presented simultaneously with A and B, respectively, and both the AX and BY compounds were paired with reward in equivalent numbers of trials. In AX+ trials, the reward was already fully predicted by the pretrained stimulus A and, therefore, the presentation of the reward should not have generated a prediction error. By contrast, in the BY+ control trials, the reward was predicted by neither stimulus, and the occurrence of reward should have generated a prediction error. The animals continued to show anticipatory licking in AX+ trials, as in the prior A+ trials, and they learned to lick in anticipation of reward in BY+ trials (Fig. 1b).

The crucial test trials with the stimulus presented alone showed that learning about stimulus X was blocked relative to the control stimulus Y which had become an effective reward predictor (Fig. 1a bottom). Median durations of licking (50th percentile) during a 2.0-s period after stimulus onset were 0 ms after stimulus X but 323 ms after Y (400 trials; $P < 0.0001$; Wilcoxon test; X and Y tested

without reward). The failure of the monkeys to respond to stimulus X in spite of its prior pairing with the reward suggests that learning depends upon the reward generating a prediction error.

An assessment of eye movements revealed that both rewarded and unrewarded stimuli were detected with comparable saccadic latencies (means in A+ and B- trials were 191 and 192 ms, respectively; $n = 133/123$; $P > 0.2$, t -test). Comparisons between stimulus and eye positions (Fig. 1c and d) demonstrate that the monkeys fixated all four stimuli despite their differential associations with reward prediction errors, including the blocked stimulus X. These data suggest that blocking was not due to a failure of the monkeys to detect stimulus X.

Conditions for neuronal learning

Based on the results from behavioural learning, we investigated how dopamine neurons acquired responses to conditioned stimuli in relation to prediction errors. We first assessed how 286 dopamine neurons discriminated between rewarded and unrewarded stimuli (8, 201 and 77 neurons in substantia nigra and ventral tegmental area groups A8, A9 and A10, respectively). We found that 200 dopamine neurons were activated by stimulus A, and that 150 of them discriminated between the reward-predicted stimulus A and nonpredictive stimulus B, either in an all-or-none fashion (75

neurons; Fig. 2a) or with weaker activations to stimulus B than A (75 neurons). Fifty of the 200 neurons showed comparable responses to the two stimuli, none of them being stronger to stimulus B than A. The incidence of nondiscriminating B responses increased slightly from 22% to 29% over several consecutively employed picture sets. Most activations to stimulus B were followed by depressions (68 of 75 neurons), thus producing characteristic biphasic responses (see below). Responding neurons were found insignificantly more frequently in the medial of three equidistant mediolateral regions of dopamine neurons ($P > 0.1$; chi-squared test). The rewarded compounds AX and BY activated 94 of 137 dopamine neurons (5, 96 and 36 neurons in groups A8, A9 and A10, respectively), none of them being activated in only one trial type (Fig. 2b). The latencies of neuronal responses were less than 100 ms and thus were shorter than the onset of saccadic eye movements towards the stimuli, indicating that the discriminations may have been based more on stimulus positions than on visual attributes requiring fixation.

Blocking was tested on responses to conditioned stimuli in 85 dopamine neurons (2, 73 and 10 neurons in groups A8, A9 and A10, respectively) (Fig. 1e). None of them showed exclusive responses to stimulus X but, more importantly, 39 were not activated by stimulus X while remaining responsive to stimulus Y (Fig. 2c). A further 16

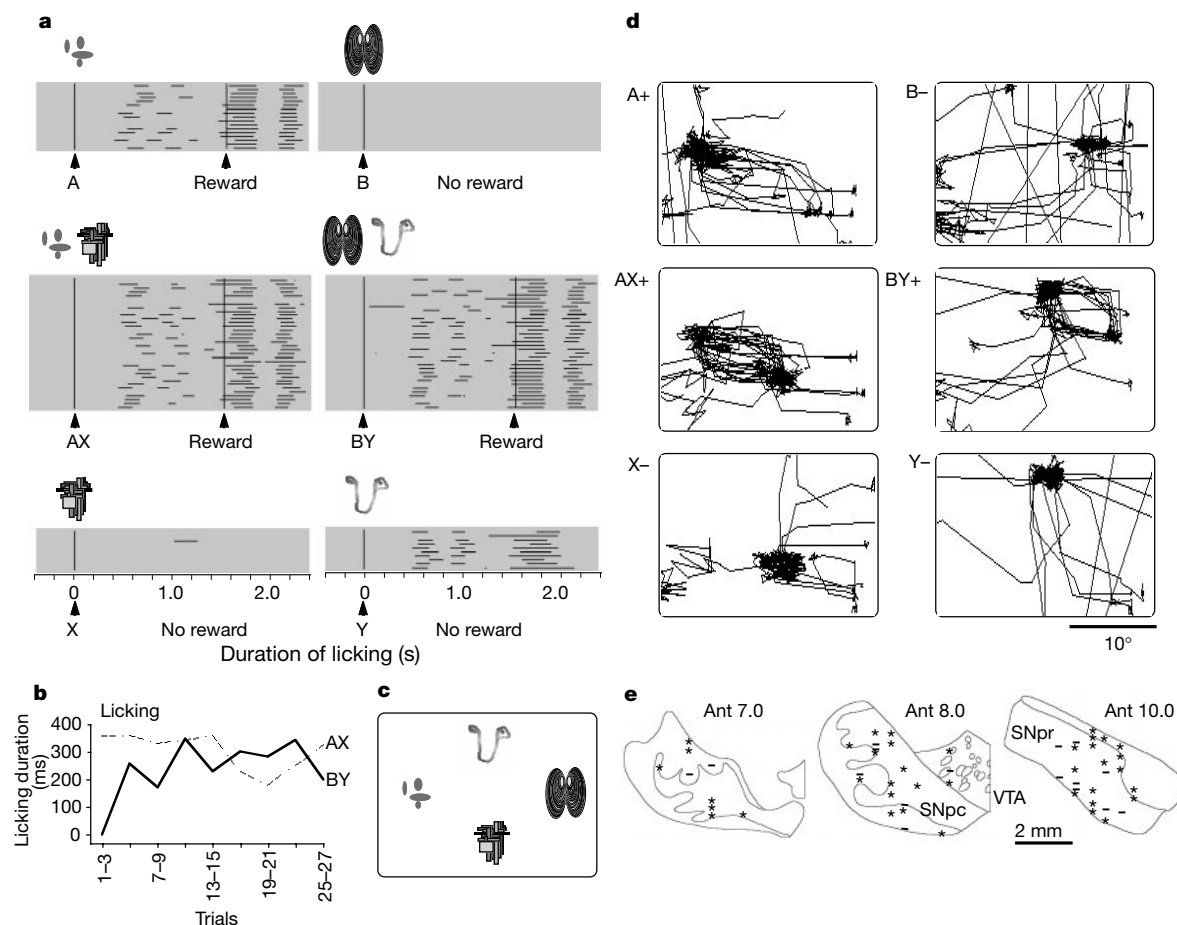


Figure 1 Behavioural performance in the blocking paradigm and neuronal localizations. **a**, Licking behaviour in the six trial types: A, B, pretraining; AX, BY, compound learning; X, Y, learning test. Horizontal lines indicate periods of licking. Note the licking in late Y-test trials in the absence of reward. **b**, Learning curves of durations of reward-anticipatory licking to compounds BY and AX during the 1.5-s stimulus-reward interval. **c**, Positions of the four stimuli on the computer screen in front of the animal. **d**, Eye positions during the 1.5-s stimulus-reward interval in the six trial types after learning (up is upward, right is rightward; eight trials are superimposed in each type). Darker areas indicate increased

eye fixations on parts of the stimuli. The eyes fixated within a 2° diameter in A+, X- and Y- trials for mean durations of 444, 367, 551 and 698 ms per trial of 1,500 ms, respectively ($F = 1.630$, degrees of freedom, d.f. = 3, $P > 0.2$, one-way ANOVA). **e**, Histologically reconstructed positions of dopamine neurons responding to stimuli X or Y (asterisks indicate exclusive activations to Y but not to X, $n = 39$; dashes indicate activations to both X and Y, $n = 16$). SNpc, substantia nigra pars compacta, SNpr, substantia nigra pars reticulata, VTA, ventral tegmental area. Ant 7.0–10.0: levels anterior to the interaural stereotaxic line.

neurons were driven by both X and Y. Of these, four showed significantly weaker activations to X than Y, 10 showed comparable activations, and two responded more strongly to X than to Y. Eight of the 16 neurons had biphasic activation–depression responses (Fig. 2d). A further 30 neurons responded to neither X nor Y. The distribution of responding neurons varied insignificantly among groups A8, A9 and A10 or among three equidistant mediolateral regions of dopamine neurons ($P > 0.3$). Thus the dopamine neurons acquired stronger responses to the reward-predicting stimulus Y than to the redundant stimulus X (Fig. 2e), although both stimuli had received the same numbers of pairings with reward during compound training. Both behavioural and neuronal learning about stimulus X was blocked in the absence of a reward prediction error.

Learning and the dopamine reward signal

Subsequently we investigated whether the acquisition of behavioural and neuronal responses to the conditioned stimuli was related to the dopamine response at the time of the reward. After pretraining, the behavioural measures indicated that the reward following stimulus A and the absence of reward following stimulus B were predicted and should not have generated prediction errors. Correspondingly, dopamine activity remained at the baseline after the reward in A trials and at the time of no reward in B trials (Fig. 3a). By contrast, a surprising reward following the usually unrewarded stimulus B should have generated a positive prediction error and did, in fact, activate 11 of 13 dopamine neurons in occasional tests (Fig. 3a bottom). Similarly, the unexpected omission of reward after stimulus A, which should generate a negative prediction error, depressed 8 of 10 dopamine neurons at the

predicted time of reward. These data suggest that the previously demonstrated dopamine reward prediction error signal^{26–28} was operational in the current paradigm.

We then assessed whether dopamine neurons differentially reported reward prediction errors during the learning of behavioural and neuronal responses. The prediction error was low during initial AX+ trials with the reward already predicted by stimulus A. Accordingly, statistically significant activations following the reward were seen in none of six dopamine neurons tested in initial blocks of AX trials alone and in only 9 of 38 dopamine neurons tested in initial AX–BY trials (Fig. 3b). By contrast, 19 of the 38 neurons showed significant reward activations in initial BY+ trials when, on the basis of pretraining with unrewarded stimulus B, the reward was unpredicted and therefore should have generated a strong prediction error. The reward activations disappeared gradually during learning and were observed as late as 234 and 610 trials after onset of compound learning with two picture sets, respectively. All six AX-tested neurons were in group A9; of the 38 neurons, 29 were in group A9 and 9 in A10. Thus the dopamine neurons showed stronger reward activations with larger positive prediction errors and weaker activations with smaller prediction errors.

The responses at the time of the reward should also reflect the differential predictive status of stimuli X and Y. As predictive learning to stimulus X was blocked, the omission of reward following this stimulus should not have generated a negative prediction error. Accordingly, only one of 85 neurons was depressed by reward omission following stimulus X, whereas 30 neurons showed a depression after reward omission following stimulus Y (Fig. 3c). The occasional test presentation of a reward after stimulus X, but

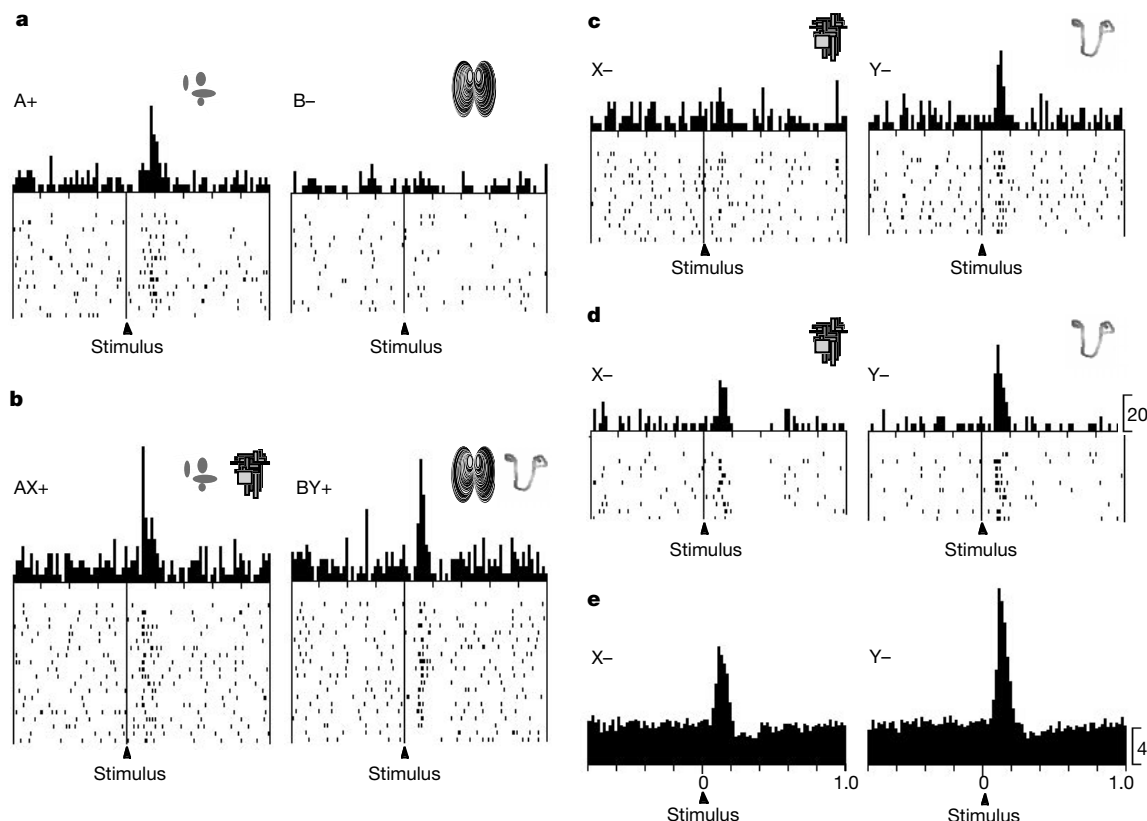


Figure 2 Acquisition of dopamine responses to conditioned stimuli depends on prediction errors in the blocking paradigm. **a**, Pretraining; **b**, after compound learning; **c–e**, learning tests. **a**, Differential activations following reward-predicting stimulus A but not unrewarded stimulus B. **b**, Maintained activation to reward-predicting compounds AX and acquired activation to BY. **c**, Absent (blocked) neuronal response to stimulus X but acquired response to stimulus Y (same neuron as in **b**). **d**, Occasional small activation

followed by depression to blocked stimulus X, probably due to stimulus generalization. **e**, Averaged population histograms of the 85 dopamine neurons tested with stimuli X and Y after compound learning (normalized for trial numbers). In **a–d**, dots denote neuronal impulses, referenced in time to the stimuli. Each line of dots shows one trial, the original sequence being from top to bottom in each panel. Histograms contain the sums of raster dots. Vertical calibrations indicate 20 impulses s^{-1} (in **a–d**) and 4 impulses s^{-1} (in **e**).

not stimulus Y, should have generated a positive prediction error and, indeed, induced an activation in all three dopamine neurons tested (Fig. 3d). Thus the contrasting activations and depressions at the time of the reward in X and Y trials corresponded to the different prediction errors generated by the presentations and omissions of the reward in these trials.

Response generalization

As noted above, 16 neurons showed activations to the blocked stimulus X, which in eight were followed by depressions (Fig. 2d). We investigated the nature of these responses by comparing them with responses to stimuli A and B, which had well defined differential reward predictions. Thirteen of the 16 X-responding neurons also showed responses to stimulus B before this stimulus had ever been paired with reward (Fig. 4a). Moreover, the strong activation by a surprising reward on occasional B test trials (Fig. 4b) confirmed that B did not predict the reward in spite of its activation of the dopamine neurons. The reward activation contrasted with the absence of any reward response following the consistently reward-predicting stimulus A (Fig. 4a). The neuronal responses to reward omission revealed similar contrasting predictive functions. Although neuronal activity was depressed by reward omission on test trials with the usually rewarded stimulus A (Fig. 4b), the activity remained at the baseline level at the time of reward on unrewarded B trials (Fig. 4a). Thus neuronal prediction errors occurred differentially at the time of the reward on A versus B trials, although both stimuli triggered neuronal activations (Fig. 4c). Therefore, the responses to stimulus B were probably not related to reward

prediction but may have reflected response generalization from the reward-predicting stimulus A (ref. 28). The notable response similarities between stimuli X and B (Figs 2d versus 4a, b) may suggest that the X responses were also unrelated to explicit reward prediction but possibly reflected generalization from the reward-predicting stimulus Y to stimulus X.

Discussion

Our data demonstrate that the activity of single dopamine neurons in the blocking paradigm conformed to basic assumptions of formal learning rules. The acquisition of neuronal responses to conditioned, reward-predicting stimuli was governed by the stringent criteria for learning developed in behavioural theories. Neuronal learning, like behavioural learning, depended crucially on the reward prediction error and occurred less readily with stimulus–reward associations alone. Furthermore, the dopamine neurons themselves coded the reward prediction error during the differential progress of behavioural and neuronal learning. Thus learning depended crucially on the presence of a reward prediction error which was coded by the dopamine neurons, and rewards that produced reduced dopamine signals failed to support both behavioural and neuronal learning. The observed concordance between cellular and behavioural processes in dopamine neurons contributes to establishing a neurobiological basis for error-driven learning rules that are derived from behavioural analyses^{2–6}. The application of formal behavioural learning rules to the study of single neurons in the mammalian brain may provide a powerful approach for studying the cellular foundations of reward-directed learning.

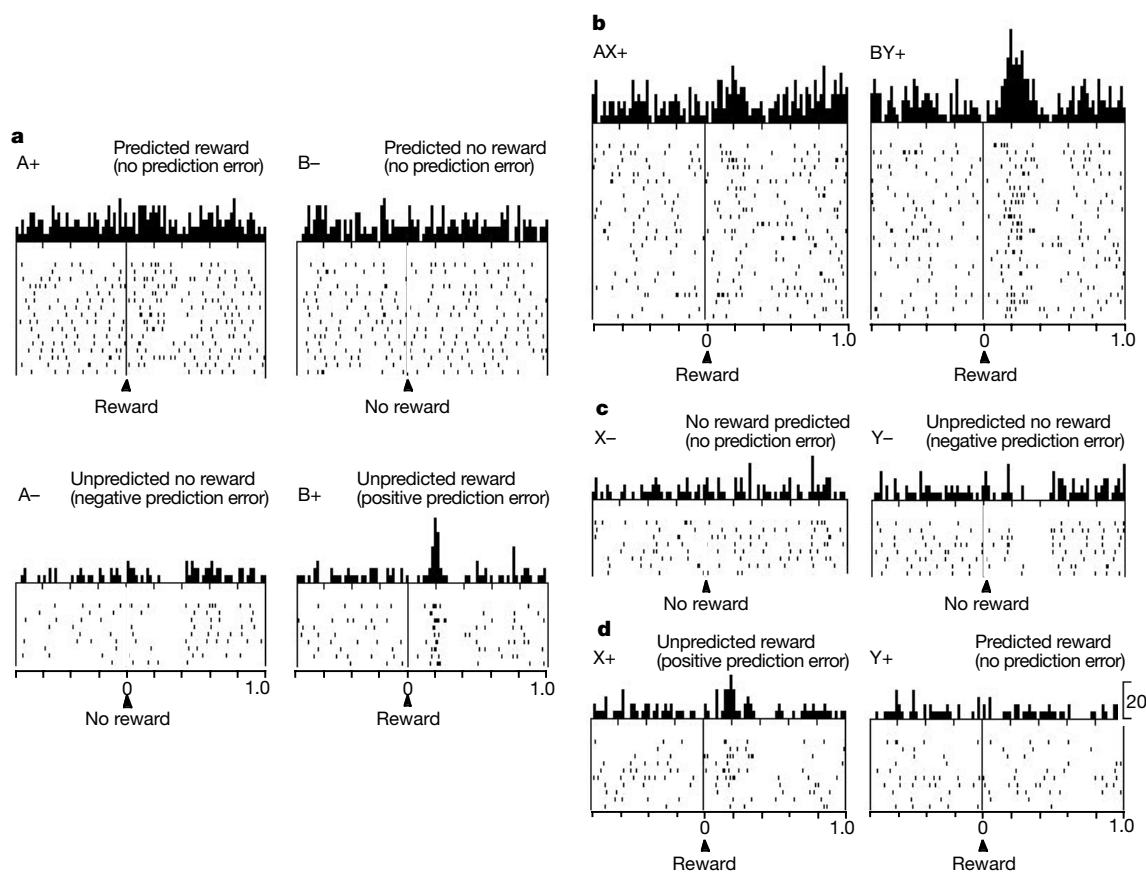


Figure 3 Dopamine prediction error response at the time of the reward in the blocking paradigm. **a**, Pretraining; **b**, during compound learning; **c**, **d**, learning tests. **a**, Lack of responses in regular trials (top); occasional test trials show neuronal depression with omitted reward in A trials and activation with surprising reward in B trials (bottom). **b**, Lack of substantial response of dopamine neuron to predicted reward in AX+ trials, but

activation to surprising reward in BY+ trials. **c**, Depressant dopamine response with omitted reward in Y but not X trials. **d**, Dopamine activation to reward in X but not Y trials. Four different neurons are shown in **a–d**, respectively. Vertical calibration indicates 20 impulses s^{-1} (in **a–d**).

The design of the blocking paradigm allowed us to differentiate the influence of reward prediction errors from that of simple stimulus–reward pairings. Although the stimuli X and Y received comparable numbers of pairings with reward during compound training, only stimulus Y was associated with a reward prediction error. The behavioural reactions demonstrated that only stimulus Y was learned as a predictor of reward, whereas X was blocked from learning. These data accord with numerous behavioural demonstrations of blocking as well as contemporary learning theory^{2–6,37,38} and suggest that behavioural learning in the present primate paradigm depends on reward prediction errors rather than on simply stimulus–reward pairings.

Dopamine neurons acquired stronger activations to stimulus Y than X, similar to the acquisition of behavioural reactions. Neuronal learning to stimulus X, like the behavioural reaction, was subject to blocking. The fewer, comparatively minor and biphasic responses to stimulus X were possibly due to response generalization, rather than neuronal reward prediction. All four stimuli used in our experiment induced ocular reactions. Thus the differential acquisition of dopamine responses to stimuli X and Y reflected the degree to which the stimuli became valid reward predictors, rather than differences in stimulus detection.

The dopamine activation at the time of the reward was stronger in BY than in AX trials. This response coincided with substantial behavioural and neuronal learning to stimulus Y. By contrast, pairings between stimulus X and the predicted reward in AX trials produced only a small dopamine response to the reward, no behavioural learning, and little acquisition of neuronal response to stimulus X. Whereas earlier work showed that dopamine reward responses diminished with increasing reward prediction during

learning²⁷, the present experiments showed a differential relationship to learning. It thus appears that the differential capacity of the reward to support behavioural and neuronal learning depends on prediction errors which are signalled by the dopamine response at the time of the reward. Rewards that produce larger reward prediction errors induce stronger dopamine activations and better behavioural and neuronal learning of a reward-predicting stimulus, thereby suggesting that dopamine reward responses and behavioural and neuronal learning are correlated.

The present results suggest that the dopamine signal encodes the prediction error of formal theories of associative learning. These theories deploy prediction errors in two distinct ways⁷. The first assumes that the error generated on a trial has a direct impact on predictive learning in that trial^{13,31}. By contrast, attentional theories argue that reward prediction errors modulate predictive learning indirectly by producing attentional learning to a stimulus that controls its subsequent associability with the reward^{4,5}. Recent neurobiological research has begun to identify the brain structures involved in such attentional learning⁴⁰. The bidirectional changes of dopamine neurons with prediction errors in A, B, X and Y trials might comply better with the $\lambda - V$ (reinforcer – predictor) error term of predictive learning than with the absolute value error term of attentional theories ($|\lambda - V|$) (ref. 7). Neurons carrying prediction errors are also found in brain structures other than the dopamine system⁷, particularly in the climbing fibre projection to cerebellar Purkinje cells^{11,12,14,15}.

The generation and action of the hypothetical dopamine teaching signal may involve basal ganglia circuits linking the dopamine neurons with modifiable synapses in the striatum and cortex^{28,35,36,41–47}. A dopamine reward-predicting teaching signal could sustain learning about extended chains of predictive stimuli leading to a primary reward by mediating the ‘credit assignment’ to earlier links of such chains through higher-order conditioning^{16,30,48}. The interpretation of the dopamine response serving as an internal reinforcement signal accords with the finding that stimuli that are blocked from learning, like dopamine responses, are weakened in their function as conditioned reinforcers to support higher-order learning⁴⁹. □

Methods

Two adult, male *Macaca fascicularis* monkeys were used in a classical conditioning procedure. Stimuli were presented for 1.5 s on a 13-inch computer monitor at 65 cm from the monkey's eyes ($24 \times 18^\circ$) (Fig. 1a, c). Each of them covered a rectangle on the screen, which had a surface area of 1,600 mm² with side lengths of 9.7 and 61.7 mm. We employed consecutively five sets of four structured, coloured visual stimuli (A, B, X, Y) and recorded from similar numbers of neurons in each set. Each neuron was tested with one stimulus set. Each stimulus set was used for several months and discarded when the next set started. Animals showed faster learning with increasing numbers of experienced stimulus sets (‘learning set’ behaviour). Stimuli were presented on a structured, coloured monitor background which was identical for each stimulus set but different between sets. In order to make the stimuli as dissimilar as possible and thus reduce potential neuronal response generalization²⁸, each stimulus had a distinctively different and constant form, mix of colours and position on the screen, and was not overlapped by other stimuli. Owing to their constant positions, animals were able to identify each stimulus most easily by its position when it came up, probably without even fixating it. In one set, stimuli X and Y were sounds delivered from a small loudspeaker below the computer monitor (2 kHz and white noise, respectively). The data obtained with the sounds were comparable to those obtained with the respective visual stimuli, including the differential responses to conditioned stimuli and rewards; they were therefore treated together. A constant 0.1–0.2-ml volume of fruit juice was delivered at the end of the 1.5-s stimulus duration through a spout at the animal's mouth. Interstimulus intervals varied randomly between 12 and 20 s. There was no specific action required by the animal for having reward delivered following a stimulus. In free liquid trial blocks, animals received 0.1–0.2 ml of fruit juice at irregular intervals of 12–20 s outside of any specific task. Lick responses of the animal were monitored by interruptions by the animal's tongue of an infrared light beam 4 mm below the spout. Eye positions were monitored through an infrared oculometer (Iscan).

Behavioural task

Three consecutive phases were employed. During pretraining, stimulus A was followed by liquid reward, but B went unrewarded. Stimuli A and B alternated semirandomly. During compound conditioning, stimulus X was added to the established, reward-predicting

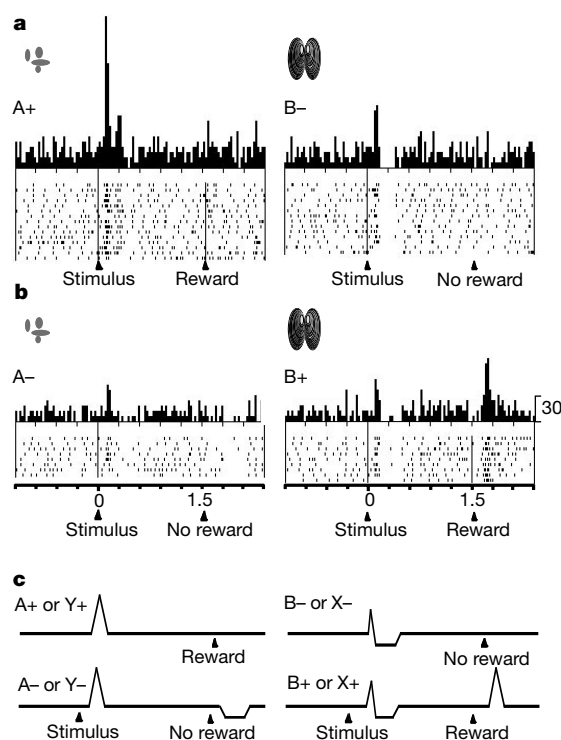


Figure 4 Dopamine responses to unrewarded stimuli may reflect stimulus generalization rather than reward prediction. **a**, Activation–depression response to unrewarded stimulus B. The absence of depression to predicted no reward in B trials indicates an absence of neuronal reward prediction. **b**, Occasional test trials in a different neuron. Activation to unpredicted reward indicates absence of reward prediction in B trials, whereas the depression at the habitual time of reward in unrewarded A trials indicates a reward prediction. Vertical calibration is 30 impulses s^{−1} for **a** and **b**. **c**, Schematic of generalizing dopamine responses to reward-predicting and non-predicting stimuli.

stimulus A without changing reward delivery. Control stimulus B was paired with stimulus Y and followed by reward. AX+ and BY+ trials were semirandomly intermixed with A+ and B- trials to maintain the reward and nonreward associations of stimuli A and B, respectively. In the third phase, stimuli X and Y were tested in occasional unrewarded trials in semirandom alternation and intermixed with A+, B-, AX+ and BY+ trials (ratio about 1:5) to avoid conditioning.

Electrophysiological recording

Activity from single midbrain dopamine neurons was recorded extracellularly during 20–60 min in the two animals using standard electrophysiological techniques. As described previously^{22–25,50} dopamine neurons discharged polyphasic, initially negative or positive impulses with relatively long durations (1.8–5.5 ms) and low frequencies (2.0–8.5 impulses s⁻¹). Impulses contrasted with those of pars reticulata neurons of substantia nigra (70–90 impulses s⁻¹ and <1.1 ms duration), a few unknown neurons discharging impulses of <1.0 ms at low rates, and neighbouring fibres (<0.4 ms duration). Only dopamine neurons activated by reward delivery in free liquid trials were tested in the present experiments (about 75–80% of the dopamine neuronal population). Neuronal activations were compared against an 800-ms control period preceding the first task event by using the Wilcoxon test in at least seven trials, with a constant time window of 70–220 ms following the conditioned stimuli and 90–220 ms following reward. These time windows comprised 80% of onset and offset times of statistically significant increases ($P < 0.01$). Neuronal depressions were assessed with the Wilcoxon test during individual time windows. Responses were compared in individual neurons with the two-tailed Mann–Whitney U test ($P < 0.05$) between different stimuli, using impulse counts in single trials. Recording sites of dopamine neurons sampled from groups A8, A9 and A10 were marked with small electrolytic lesions and reconstructed from 40- μ m-thick, tyrosine-hydroxylase-immunoreacted or cresyl-violet-stained, stereotactically oriented, coronal brain sections (Fig. 1e). Experimental protocols conformed to the Swiss Animal Protection Law and were supervised by the Fribourg Cantonal Veterinary Office.

Received 8 February; accepted 17 May 2001.

1. Thorndike, E. L. *Animal Intelligence: Experimental Studies* (MacMillan, New York, 1911).
2. Pavlov, I. P. *Conditional Reflexes* (Oxford Univ. Press, London, 1927).
3. Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. H. & Prokasy, W. F.) 64–99 (Appleton Century Crofts, New York, 1972).
4. Mackintosh, N. J. A theory of attention: Variations in the associability of stimulus with reinforcement. *Psychol. Rev.* **82**, 276–298 (1975).
5. Pearce, J. M. & Hall, G. A. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
6. Dickinson, A. *Contemporary Animal Learning Theory* (Cambridge Univ. Press, Cambridge, 1980).
7. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
8. Widrow, G. & Hoff, M. E. Adaptive switching circuits. *IRE Western Electron. Show Convention, Convention Record Part 4*, 96–104 (1960).
9. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng. Trans. ASME* **82**, 35–45 (1960).
10. Widrow, G. & Sterns, S. D. *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, 1985).
11. Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202**, 437–470 (1969).
12. Ito, M. Long-term depression. *Ann. Rev. Neurosci.* **12**, 85–102 (1989).
13. Thompson, R. F. & Gluck, M. A. in *Perspectives on Cognitive Neuroscience* (eds Lister, R. G. & Weingartner, H.) 25–45 (Oxford Univ. Press, New York, 1991).
14. Kawato, M. & Gomi, H. The cerebellum and VOR/OKR learning models. *Trends Neurosci.* **15**, 445–453 (1992).
15. Kim, J. J., Krupa, D. J. & Thompson, R. F. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science* **279**, 570–573 (1998).
16. Sutton, R. S. & Barto, A. G. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* **88**, 135–170 (1981).
17. Sutton, R. S. & Barto, A. G. *Reinforcement Learning* (MIT Press, Cambridge, Massachusetts, 1998).
18. Fibiger, H. C. & Phillips, A. G. in *Handbook of Physiology—The Nervous System IV* (ed. Bloom, F. E.) 647–675 (Williams and Wilkins, Baltimore, 1986).
19. Wise, R. A. & Hoffman, C. D. Localization of drug reward mechanisms by intracranial injections. *Synapse* **10**, 247–263 (1992).
20. Robinson, T. E. & Berridge, K. C. The neural basis for drug craving: an incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).
21. Robbins, T. W. & Everitt, B. J. Neurobehavioural mechanisms of reward and motivation. *Curr. Opin. Neurobiol.* **6**, 228–236 (1996).

22. Romo, R. & Schultz, W. Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* **63**, 592–606 (1990).
23. Schultz, W. & Romo, R. Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioural reactions. *J. Neurophysiol.* **63**, 607–624 (1990).
24. Ljungberg, T., Apicella, P. & Schultz, W. Responses of monkey dopamine neurons during learning of behavioural reactions. *J. Neurophysiol.* **67**, 145–163 (1992).
25. Schultz, W., Apicella, P. & Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913 (1993).
26. Schultz, W., Dayan, P. & Montague, R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
27. Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* **1**, 304–309 (1998).
28. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
29. Salamone, J. D. The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav. Brain Res.* **61**, 117–133 (1994).
30. Horvitz, J. C. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience* **96**, 651–656 (2000).
31. Sutton, R. S. & Barto, A. G. in *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel, M. & Moore, J.) 497–537 (MIT Press, Cambridge, Massachusetts, 1990).
32. Mackintosh, N. J. *Conditioning and Associative Learning* (Oxford Univ. Press, New York, 1983).
33. Friston, K. J., Tononi, G., Reeke, G. N. Jr, Sporns, O. & Edelman, G. M. Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience* **59**, 229–243 (1994).
34. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
35. Houk, J. C., Adams, J. L. & Barto, A. G. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J. C., Davis, J. L. & Beiser, D. G.) 249–270 (MIT Press, Cambridge, Massachusetts, 1995).
36. Suri, R. & Schultz, W. A neural network with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* **91**, 871–890 (1999).
37. Kamin, L. J. in *Fundamental Issues in Instrumental Learning* (eds Mackintosh, N. J. & Honig, W. K.) 42–64 (Dalhousie Univ. Press, Dalhousie, 1969).
38. Martin, I. & Levey, A. B. Blocking observed in human eyelid conditioning. *Q. J. Exp. Psychol. B* **43**, 233–255 (1991).
39. Dickinson, A. Causal learning: An associative analysis. *Q. J. Exp. Psychol. B* **54**, 3–25 (2001).
40. Holland, P. C. Brain mechanisms for changes in processing of conditioned stimuli in Pavlovian conditioning: Implications for behavioural theory. *Anim. Behav.* **25**, 373–399 (1997).
41. Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B. & Bernardi, G. Long-term synaptic depression in the striatum: Physiological and pharmacological characterization. *J. Neurosci.* **12**, 4224–4233 (1992).
42. Garcia-Munoz, M., Young, S. J. & Groves, P. Presynaptic long-term changes in excitability of the corticostriatal pathway. *NeuroReport* **3**, 357–360 (1992).
43. Wickens, J. R., Begg, A. J. & Arbuthnott, G. W. Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* **70**, 1–5 (1996).
44. Calabresi, P. et al. Abnormal synaptic plasticity in the striatum of mice lacking dopamine D2 receptors. *J. Neurosci.* **17**, 4536–4544 (1997).
45. Otani, S., Blond, O., Desce, J. M. & Crepel, F. Dopamine facilitates long-term depression of glutamatergic transmission in rat prefrontal cortex. *Neuroscience* **85**, 669–676 (1998).
46. Otani, S., Auclair, N., Desce, J., Roisin, M. P. & Crepel, F. *J. Neurosci.* **19**, 9788–9802 (1999).
47. Centonze, D. et al. Unilateral dopamine denervation blocks corticostriatal LTP. *J. Neurophysiol.* **82**, 3575–3579 (1999).
48. Minsky, M. L. Steps toward artificial intelligence. *Proc. Inst. Radio Engineers* **49**, 8–30 (1961).
49. Rauhut, A. S., McPhee, J. E. & Ayres, J. B. Blocked and overshadowed stimuli are weakened in their ability to serve as blockers and second-order reinforcers in Pavlovian fear conditioning. *J. Exp. Psychol. Anim. Behav. Process* **25**, 45–67 (1999).
50. Schultz, W. & Romo, R. Responses of nigrostriatal dopamine neurons to high intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* **57**, 201–217 (1987).

Acknowledgements

We thank B. Aebischer, J. Corpataux, A. Gaillard, B. Morandi, A. Pisani and F. Tinguely for expert technical assistance. The study was supported by the Swiss NSF, the European Union (Human Capital and Mobility, and Biomed 2 programmes), the James S. McDonnell Foundation and the British Council.

Correspondence and requests for materials should be addressed to W.S. (e-mail: Wolfram.Schultz@unifr.ch).

A giant stream of metal-rich stars in the halo of the galaxy M31

Rodrigo Ibata*, Michael Irwin†, Geraint Lewis‡, Annette M. N. Ferguson§ & Nial Tanvir||

* Observatoire de Strasbourg, 11, rue de l'Université, F-67000 Strasbourg, France

† Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

‡ Anglo-Australian Observatory, PO Box 296, Epping, New South Wales 1710, Australia

§ Kapteyn Astronomical Institute, Postbus 800, 9700 AV Groningen, The Netherlands

|| Physical Sciences, University of Hertfordshire, Hatfield AL10 9AB, UK

Recent observations have revealed streams of gas and stars in the halo of the Milky Way^{1–3} that are the debris from interactions between our Galaxy and some of its dwarf companion galaxies; the Sagittarius dwarf galaxy and the Magellanic clouds. Analysis of the material has shown that much of the halo is made up of cannibalized satellite galaxies^{2,4}, and that dark matter is distributed nearly spherically in the Milky Way. It remains unclear, however, whether cannibalized substructures are as common in the haloes of galaxies as predicted by galaxy-formation theory⁵. Here we report the discovery of a giant stream of metal-rich stars within the halo of the nearest large galaxy, M31 (the Andromeda galaxy). The source of this stream could be the dwarf galaxies M32 and NGC205, which are close companions of M31 and which may have lost a substantial number of stars owing to tidal interactions. The results demonstrate that the epoch of galaxy building still continues, albeit at a modest rate, and that tidal streams may be a generic feature of galaxy haloes.

Within the framework of hierarchical structure formation, large spiral galaxies like the Milky Way or Andromeda arose from the merger of many small galaxies and protogalaxies⁶. Later in their evolution, spiral galaxies become the dominant component in such mergers, cannibalizing smaller systems that fall within their sphere of influence. The complete destruction of the victim is usually progressive, and may take several orbits. However, the stellar debris from the destroyed dwarf galaxy follows a similar orbital trajectory to the progenitor, which is likely to have started life far away from the place of its final demise, and so the tidally disrupted matter tends to be deposited over a broad range in distance from the larger galaxy. Over time, with the accumulation of many such mergers, large galaxies develop an extensive stellar and dark-matter 'halo', the latter being by far the most massive component of the galaxy. Meanwhile, part of the (dissipative) gas component of the smaller galaxies feeds the growth of the disk of the larger galaxy. This is seen in numerical simulations of galaxy formation, which result in galactic haloes comprising clumps of dark matter⁵. If this prediction is correct, then haloes should possess significant substructure—in contrast to previous suggestions⁷, which predict the dark and luminous components of haloes to be distributed smoothly.

Most of the halo of the Milky Way is metal-poor and, to first order, smoothly distributed; but recent studies have shown that the halo contains non-negligible stellar substructure. Evidence for the phase-space clumping of halo stars has even been found in the solar neighbourhood, where about 10% of the stars may be associated with a single ancient accretion event⁴. The discovery that the Milky Way is surrounded by a giant rosette-like stream originating from the Sagittarius dwarf galaxy^{2,3} shows that—on the largest scales—the structure of the halo is substantially 'streamy': approximately half of the intermediate-age stars at distances beyond about 15 kpc belong to the Sagittarius stream. This also implies² that the last large accretion was that of the Sagittarius dwarf, and that the Milky Way

has not cannibalized many other small galaxies for about 7 Gyr. Unless there has been a continual accretion of highly dark-matter-dominated 'galaxies', or of galaxies containing exclusively old stars, the formation of the Galactic halo must have been essentially complete at a point in time less than half the age of the Universe.

We now consider whether the behaviour of the Milky Way is unusual. The only external galaxy in which halo substructure has been reported is NGC5907, which possesses a gigantic extraplanar stellar stream⁸, substantially brighter than similar structures in the Milky Way. However, this galaxy has long been known to be peculiar⁹, having a red, and luminous, flattened 'halo', probably the result of the strong interaction that deposited the stream. To understand whether stream-like substructure is the generic morphology of galaxy haloes, it is necessary to investigate other 'normal' galaxies like the Milky Way. The prime target for such a study is the Milky Way's 'sister' galaxy, M31—the Andromeda nebula, which at a distance of ~780 kpc is the closest large galaxy¹⁰.

The Wide Field camera¹¹ on the 2.5-m Isaac Newton Telescope (INT WFC) is a four-chip charge-coupled device (CCD) mosaic camera imaging about 0.3 degree² per exposure. On the nights of 3–9 September 2000, this instrument was used to survey the southeastern half of the halo of the Andromeda galaxy. To tile this region out to a distance of 4° (~55 kpc in projection) from the centre of M31 required 58 contiguous fields. Images were taken in the equivalent of Johnson visual V and Gunn i bands under good atmospheric conditions, with 85% of fields taken in photometric conditions with seeing better than 1.2".

Previous imaging studies of the M31 halo and outer disk have either sampled the outer parts of M31 at only a few discrete locations^{12–14}, or have taken a panoramic, but much shallower, view^{15,16}. In contrast, our deep survey allows us to make an uninterrupted study of the spatial variations in stellar density as a function of magnitude and colour over a large fraction of the halo. The continuous nature of this survey enables us to distinguish local density enhancements from both the large-scale structure of the halo of M31 and the underlying foreground Galactic distribution of stars. The exposure time of 800–1,000 s per passband per field reaches i-band magnitude $i = 23.5$ and V-band magnitude $V = 24.5$ (with a signal-to-noise ratio of ~5) and allows detection of individual red-giant branch (RGB) stars to an absolute V-band magnitude $M_V = 0$ and main-sequence stars to $M_V = -1$ in the halo of M31.

The WFC pipeline processing provides internal cross-calibration for the four INT WFC CCDs (each of 4,096 × 2,048 pixels) at a level of about 1% within each pointing. Field-to-field variations in photometric zero points were calibrated and cross-checked using a combination of multiple nightly photometric standard sequence observations and the overlap regions between adjacent WFC pointings. The overall derived photometric zero-points for the whole survey are good to the level of 1–2% in both bands. (Details of the data processing and calibration will be presented elsewhere.) Objects were classified as noise artefacts, galaxies or stars according to their morphological structure on all the images.

Figure 1 presents a summary of the results of our survey. Visual inspection of the surface density of sources classified as star-like on the i-band images and with magnitudes and colours consistent with the known properties of RGB stars at the distance of M31, shows the presence of a stream-shaped over-density of sources in the halo close to, but distinct from, the minor axis of M31. The RGB stellar density in the halo increases on average by a factor of two in the on-stream regions, and is statistically significant at ~50σ level. For stars brighter than the tip of the RGB in M31, the spatial density of sources is smooth and shows no sign of the feature: the stream therefore cannot be a foreground Galactic population. The on-stream stars follow a similar colour–magnitude sequence to the RGB stars in the remainder of the halo of M31, but with evidence of an enhanced metallicity relative to the 'normal' M31 halo popula-

tion owing to their redder colours. The average V-band surface brightness of the stream is $\Sigma_V \approx 30 \pm 0.5 \text{ mag arcsec}^{-2}$, and the stream extends out to the current limit of our survey, at a projected distance of about 40 kpc.

The metallicity of the stream, deduced from the colour–magnitude diagrams displayed in Fig. 2, covers a broad range with a mean slightly more metal-rich than the Galactic globular cluster 47 Tucanae (whose metallicity, that is, whose ratio of iron to hydrogen compared to that of the Sun, is $[\text{Fe}/\text{H}] = -0.7$). We note that stars of near-solar metallicity are concentrated in the stream. Our survey reveals that such high-metallicity stars are also sparsely distributed throughout large parts of the halo of M31, consistent with the results of previous studies^{12,17}. This high metallicity, with a mean approximately 10 times greater than that of the Milky Way's halo, and the overly large stellar density of the halo—a factor of 10

greater¹⁴ than that of the Galactic halo—have until now been a puzzle.

What is this stream-like feature? The outer regions of galactic disks are metal-poor¹⁸, so the metallicity of the feature argues against a direct association with the outer disk of Andromeda. Furthermore, if it were a disk structure, it would be located at an impossible de-projected radial distance of about 140 kpc from the centre of the galaxy. This distance constraint is weakened if the feature is part of a pronounced disk warp, but that possibility is highly unlikely given the observed elongated structure. The only plausible explanation is that it is part of a large stellar stream within the halo. This stream lies along a line connecting the Andromeda satellites M32 and NGC205, and is aligned with the direction of elongation of the outer isophotes of NGC205, suggesting a relationship between the Andromeda stream and these two dwarf galaxies.

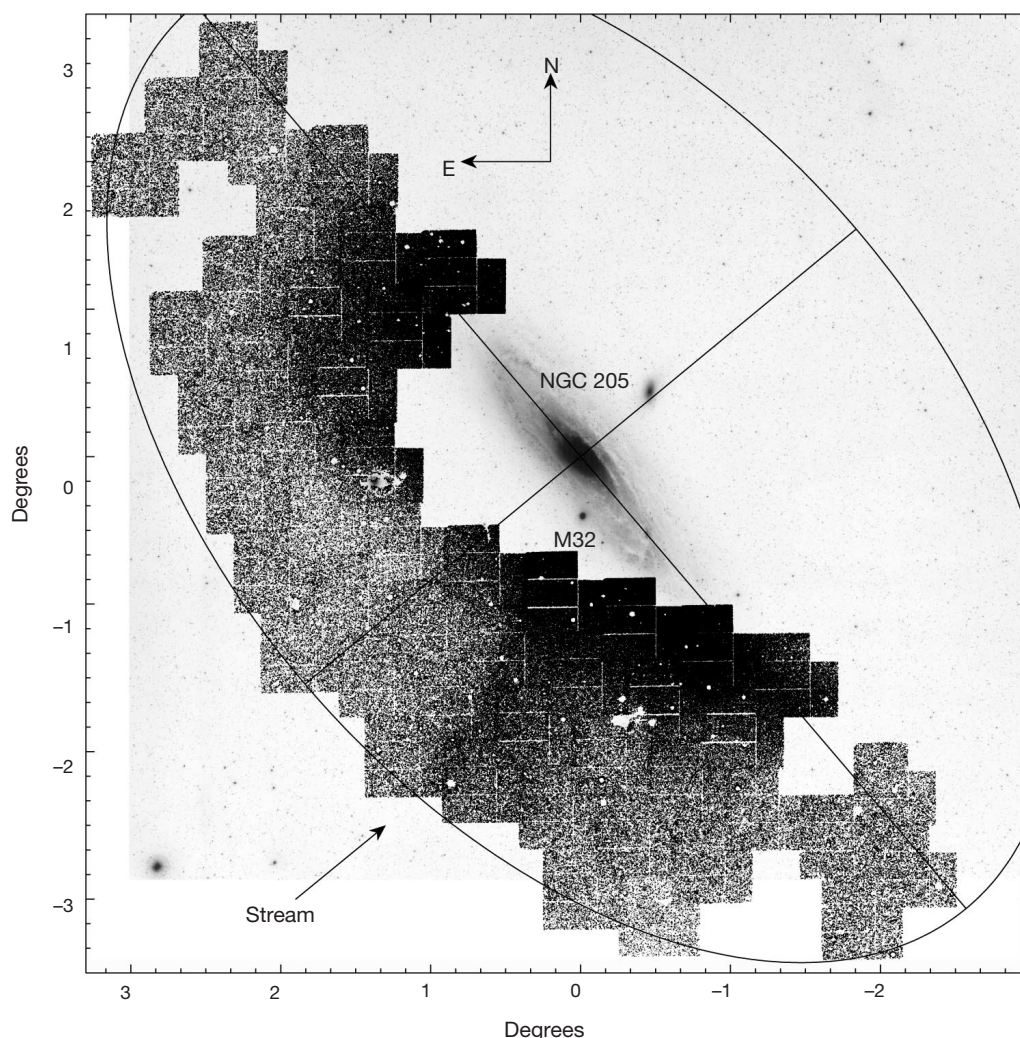


Figure 1 Surface density of red-giant branch (RGB) stars over the outer southeastern halo of M31. The surface-density maps produced by our survey are shown as 'tiles' overlying an optical image centred on M31 (see below). The over-density of stars revealed in the present study is seen as a stream extending out of M31 close to its minor axis. The density of extended background galaxies (approximately 25,000 per degree² to $V < 24$ mag) becomes comparable to the stellar density at a projected distance of 20–25 kpc from the centre of M31 along the minor axis. A fraction of these galaxies are misclassified by the reduction algorithm, thereby contaminating the stellar sample. There is a further contamination from foreground Galactic disk dwarfs, which number $<10\%$ of the M31 halo population over the survey region. However, the contribution of these contaminating sources to the number density maps is easily removed. Indeed, the stream can be

detected at signal-to-noise ratio >5 over tiny individual regions of area $>0.01 \text{ degree}^2$. The gaps to the northeast and southwest of the map correspond to fields taken in poor seeing conditions where the image quality criteria of the survey were not attained. The surface density maps are superimposed upon a photographic Palomar Sky Survey image of M31; clearly visible are the two companion galaxies M32 and NGC 205. The major and minor axes of M31 are displayed, and an ellipse denoting the size of a flattened ellipsoid (aspect ratio 3:5) of semi-major axis length 55 kpc has been superimposed to show the spatial extent of the survey. The location of the Andromeda stream is labelled. The overall shape of the M31 halo revealed by our observations appears rather boxy, with indications of other possible substructures—but these results require further study to confirm their reality.

The total absolute magnitude of the stellar stream is $M_V \approx -14$, which we estimate from the differential luminosity function between equivalent on- and off-stream fields (plausible assumptions were made to allow for fainter stream stars, the incompleteness of the survey and the uniformity of the stream). This luminosity is a factor of about 10 lower than that of either M32 and NGC205, consistent with the possibility that the stream is the debris stripped from one (or both) of these two dwarf galaxies during a recent interaction with M31.

M32 and NGC205 are both unusual dwarf elliptical galaxies which lie at projected distances of, respectively, only 5 kpc and 9 kpc from the centre of M31. NGC205 is still active in forming stars, and has had a varied and complex star-formation history¹⁹. The central regions are also known to contain dust, H I gas and molecular gas²⁰, and there is substantial morphological²¹ and kinematic²² evidence that indicates that NGC205 is being tidally distorted and potentially disrupted. The H I distribution shows a well defined velocity gradient (unlike the stars), and is less extended than the overall stellar content of the galaxy^{20,23}. The clear inconsistency between the gas and stars suggests that the gas may have recently been captured, possibly from the disk of M31.

M32, on the other hand, does not appear to be substantially distorted, yet it does seem to have a significant population of intermediate-age stars in addition to a classical old stellar component^{24,25}. There is also clear evidence of a large metallicity spread in the giant branch, with a mean just below solar, supporting a prolonged star-forming epoch²⁶.

The broad agreement of the metallicity distributions of the stream stars and these two dwarf satellites, together with their alignment and physical proximity to M31, point to a common origin. Furthermore, the unusual properties of most of the rest of the halo of M31 is also consistent with its relatively recent tidal

origin. It seems quite likely that the apparently peculiar properties of M31's halo are simply the result of a prolonged, aggressive bout of tidal stripping from either one, or both, of its two nearest neighbour satellite galaxies.

If this interpretation is correct, the stream (and possibly most of the stellar halo) has to be the result of previous interactions with M31, as there is otherwise not enough time to spatially separate if from either of the two dwarf galaxies. Furthermore, by comparison with numerical models of the Sagittarius stream^{27,28}, there will be a similar stream (either leading or trailing the progenitor dwarf(s), depending on the sense of their orbit) on the opposite side of M31.

During disk-crossing episodes, gas in the disk of M31 and in these dwarf galaxies is shocked, leading to episodic bursts of star formation, and gas exchange. In the dwarf galaxies, this gas may either be recycled from earlier generations of stars or accreted from the M31 disk. These complex interactions may explain the unusual star-formation history of the dwarf galaxies and the Andromeda stream. In turn, the regular impacts perturb the disk of M31, possibly enough to induce its warped shape²⁹ (it is plausible that a similar process is responsible for the warp in the disk of the Milky Way³⁰).

The discovery of the Andromeda stream in the first deep, panoramic survey of the Milky Way's nearest large companion suggests that halo substructure in the form of giant tidal streams may be a generic property of large spiral galaxies, and that the formation of galaxies continues at a moderate pace up to the present day. Although the dwarf galaxies M32 and NGC205 may be the source of this material, we cannot rule out the alternative possibility that the stream may be the fossil remnant of a third cannibalized system, found at a stage intermediate to those seen within the Milky Way^{2,4}.

To understand more fully the origin and evolution of the

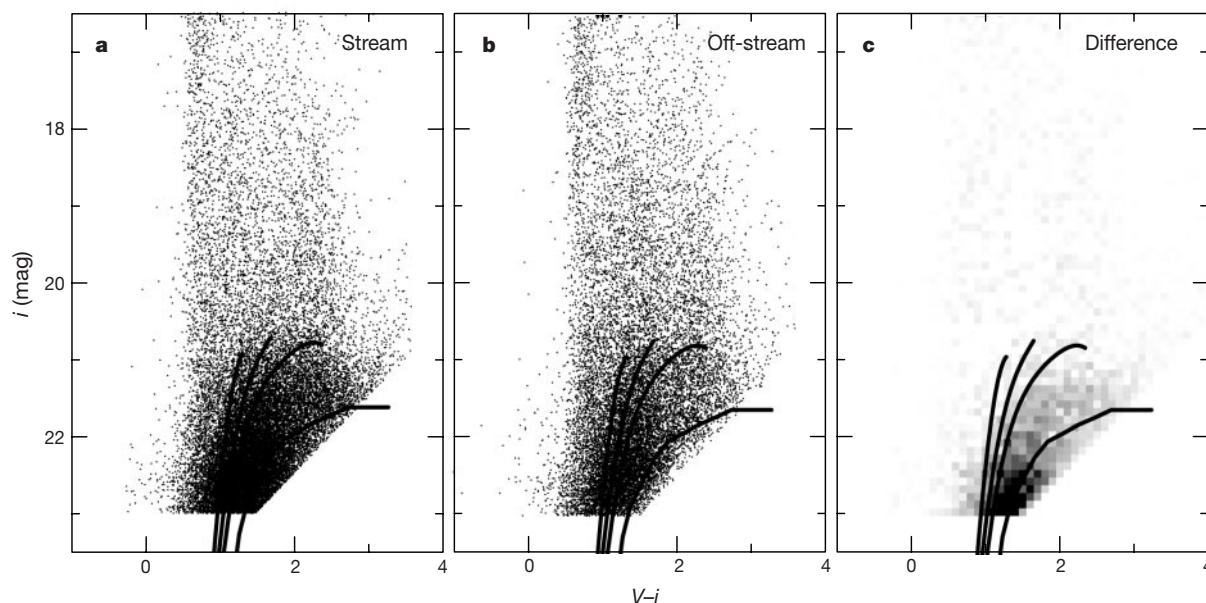


Figure 2 The relation between $V-i$ colour and i -band magnitude. Data are shown for a stream field (**a**), and an off-stream field (**b**). **a** and **b** both cover 0.3 degree^2 of sky, and were taken in similar observing conditions ($0.97''$ in i and $0.93''$ in V for the stream field, and $1.08''$ for both i and V for the off-stream field). **a** and **b** are also at a similar distance ($\sim 2''$) from the centre of M31, with the off-stream field almost exactly on the minor axis of the galaxy. The red-giant branch (RGB) stars of M31 are seen at magnitudes fainter than $i = 20.5$. To guide the interpretation of these diagrams, we have superimposed on each panel the RGB tracks (shifted to the distance of M31 and corrected for extinction) of four well-studied galactic globular clusters of different metallicity ($[\text{Fe}/\text{H}]$); these are, from left to right, NGC6397 ($[\text{Fe}/\text{H}] = -1.91$), NGC1851 ($[\text{Fe}/\text{H}] = -1.29$), 47 Tuc

($[\text{Fe}/\text{H}] = -1.71$) and NGC6553 ($[\text{Fe}/\text{H}] = -0.2$). The difference between the stellar populations in **a** and **b** is displayed in **c** in the form of a binned colour-magnitude diagram. With much of the contamination removed in this way, we see that the stream contains a broad range of stellar populations, from metallicities similar to that of NGC1851 through to solar abundance, with a mean metallicity slightly higher than that of 47 Tuc. It is particularly difficult to constrain the distance to the stream owing to this metallicity spread; to within an uncertainty of possibly as much as 0.5 mag the stream RGB appears at the same position as the M31 RGB, which suggests a distance of $d \approx 800 \pm 200 \text{ kpc}$. (The photometric uncertainties are approximately 0.2 mag at the faint limit of the data in the diagrams; 0.1 mag uncertainties occur at $V = 23.7 \text{ mag}$, $i = 22.5 \text{ mag}$).

Andromeda stream, a number of follow-up observational programmes are required, including an extension of the panoramic survey not only to the northern regions of the halo of M31, but also out to larger radii, as it is evident in the present data that the stream extends beyond the 40-kpc limits of the survey. The proximity of M31 also provides us with an opportunity to undertake a spectroscopic survey of individual stars within the stream, allowing us to map its kinematic and chemical properties. As with the tidal material detected in the halo of our own Galaxy, studies of the Andromeda stream would allow us to map the distribution of dark matter within the halo of our nearest neighbouring galaxy, as well as furthering our understanding of the process of galaxy formation. □

Received 9 February; accepted 4 May 2001.

- Putman, M. E. *et al.* Tidal disruption of the Magellanic Clouds by the Milky Way. *Nature* **394**, 752–754 (1998).
- Ibata, R., Lewis, G. F., Irwin, M., Totten, E. & Quinn, T. Great circle tidal streams: evidence for a nearly spherical massive dark halo around the Milky Way. *Astrophys. J.* **551**, 294–311 (2001).
- Ibata, R., Irwin, M., Lewis, G. F. & Stolte, A. Galactic halo substructure in the Sloan Digital Sky Survey: The ancient tidal stream from the Sagittarius dwarf galaxy. *Astrophys. J.* **547**, L133–L136 (2001).
- Helmi, A., White, S. D. M., de Zeeuw, P. T. & Zhao, H. Debris streams in the solar neighbourhood as relics from the formation of the Milky Way. *Nature* **402**, 53–55 (1999).
- Klyin, A., Gottlöber, S., Kravtsov, A. V. & Khokhlov, A. M. Galaxies in N-body simulations: Overcoming the overmerging problem. *Astrophys. J.* **516**, 530–551 (1999).
- Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F. & Zepf, S. E. A recipe for galaxy formation. *Mon. Not. R. Astron. Soc.* **271**, 781–806 (1994).
- Eggen, O. J., Lynden-Bell, D. & Sandage, A. R. Evidence from the motions of old stars that the Galaxy collapsed. *Astrophys. J.* **136**, 748–766 (1962).
- Shang, Z. *et al.* Ring structure and warp of NGC 5907: Interaction with dwarf galaxies. *Astrophys. J.* **504**, L23–L26 (1998).
- Sackett, P. D., Morrison, H. L., Harding, P. & Boroson, T. A. A faint luminous halo that may trace the dark matter around spiral galaxy NGC 5907. *Nature* **370**, 441–443 (1994).
- Stanek, K. Z. & Garnavich, P. M. Distance to M31 with the Hubble Space Telescope and HIPPARCOS red clump stars. *Astrophys. J.* **503**, L131–L134 (1998).
- Irwin, M. & Lewis, J. INT WFS pipeline processing. *New Astron. Rev.* **45**, 105–110 (2001).
- Holland, S., Fahlman, G. G. & Richer, H. B. Deep HST V- and I-band observations of the halo of M31: Evidence for multiple stellar populations. *Astron. J.* **112**, 1035–1045 (1996).
- Rich, R. M., Mighell, K. J., Freedman, W. L. & Neill, J. D. Local Group populations with the Hubble Space Telescope. I. The M31 globular cluster G1–Mayall II. *Astron. J.* **111**, 768–776 (1996).
- Reitzel, D. B., Guhathakurta, P. & Gould, A. Isolating red giant stars in M31's elusive outer spheroid. *Astron. J.* **116**, 707–722 (1998).
- Walterbos, R. A. M. & Kennicutt, R. C. Multi-color photographic surface photometry of the Andromeda galaxy. *Astron. Astrophys. Suppl.* **69**, 311–332 (1987).
- Innanen, K. A., Kamper, K. W., van den Bergh, S. & Papp, K. A. The optical warp of M31. *Astrophys. J.* **254**, 515–516 (1982).
- Durrell, P. R., Harris, W. E., Pritchett, C. J. & Davidge, T. Photometry of the outer halo of M31. *Am. Astron. Soc. Meeting* **195**, 4.04 (1999).
- Ferguson, A. M. N., Gallagher, J. S. & Wyse, R. F. G. The extreme outer regions of disk galaxies. I. Chemical abundances of H II regions. *Astron. J.* **116**, 673–690 (1998).
- Lee, M. G. Stellar population is the central region of the dwarf elliptical galaxy NGC 205. *Astron. J.* **112**, 1438–1449 (1996).
- Young, L. M. & Lo, K. Y. The neutral interstellar medium in nearby dwarf galaxies. II. NGC 185, NGC 205, and NGC 147. *Astrophys. J.* **476**, 127–143 (1997).
- Hodge, P. W. The structure and content of NGC 205. *Astrophys. J.* **182**, 671–696 (1973).
- Bender, R., Paquet, A. & Nieto, J. Internal stellar kinematics of three dwarf ellipticals in the Local Group. *Astron. Astrophys.* **246**, 349–353 (1991).
- Welch, G. A., Sage, L. J. & Mitchell, G. F. The puzzling features of the interstellar medium in NGC 205. *Astrophys. J.* **499**, 209–220 (1998).
- Freedman, W. L. Stellar content of nearby galaxies. II—the Local Group dwarf elliptical galaxy M32. *Astron. J.* **98**, 1285–1304 (1989).
- Davidge, T. J. The evolved red stellar content of M32. *Publ. Astron. Soc. Pacif.* **112**, 1177–1187 (2000).
- Grillmair, C. J. *et al.* Hubble Space Telescope observations of M32: The color-magnitude diagram. *Astron. J.* **112**, 1975–1987 (1996).
- Johnston, K. V., Spergel, D. N. & Hernquist, L. The disruption of the Sagittarius dwarf galaxy. *Astrophys. J.* **451**, 598–606 (1995).
- Ibata, R. A. & Lewis, G. F. Galactic indigestion: Numerical simulations of the Milky Way's closest neighbor. *Astrophys. J.* **500**, 575–590 (1998).
- Schwarzschild, M. Mass distribution and mass-luminosity ratio in galaxies. *Astron. J.* **59**, 273–284 (1954).
- Ibata, R. A. & Razoumov, A. O. Archer of the Galactic disk? The effect on the outer HI disk of the Milky Way of collisional encounters with the Sagittarius dwarf galaxy. *Astron. Astrophys.* **336**, 130–136 (1998).

Acknowledgements

This paper is based on observations made with the Isaac Newton Telescope operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias.

Correspondence and requests for materials should be addressed to R.I. (e-mail: ibata@astro.u-strasbg.fr).

Dynamical tunnelling of ultracold atoms

W. K. Hensinger^{*†}, H. Häffner^{*}, A. Browaeys^{*}, N. R. Heckenberg[†], K. Helmerson^{*}, C. McKenzie^{*}, G. J. Milburn[‡], W. D. Phillips^{*}, S. L. Rolston^{*}, H. Rubinsztein-Dunlop[†] & B. Upcroft[†]

^{*} National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

[†] Centre for Laser Science, Department of Physics, [‡] Centre for Quantum Computer Technology, The University of Queensland, Brisbane, Queensland 4072, Australia

The divergence of quantum and classical descriptions of particle motion is clearly apparent in quantum tunnelling^{1,2} between two regions of classically stable motion. An archetype of such non-classical motion is tunnelling through an energy barrier. In the 1980s, a new process, ‘dynamical’ tunnelling^{1–3}, was predicted, involving no potential energy barrier; however, a constant of the motion (other than energy) still forbids classically the quantum-allowed motion. This process should occur, for example, in periodically driven, nonlinear hamiltonian systems with one degree of freedom^{4–6}. Such systems may be chaotic, consisting of regions in phase space of stable, regular motion embedded in a sea of chaos. Previous studies predicted⁴ dynamical tunnelling between these stable regions. Here we observe dynamical tunnelling of ultracold atoms from a Bose–Einstein condensate in an amplitude-modulated optical standing wave. Atoms coherently tunnel back and forth between their initial state of oscillatory motion (corresponding to an island of regular motion) and the state oscillating 180° out of phase with the initial state.

Tunnelling between discrete states is coherent and reversible—that is, the particle oscillates between the states, which is the process we describe here. Coherent tunnelling oscillations are especially significant, as the system must pass through states corresponding to superpositions of distinct classical motions. It has been suggested⁷ that such states are strongly suppressed by decoherence, a subject of central importance to quantum information. Our system should

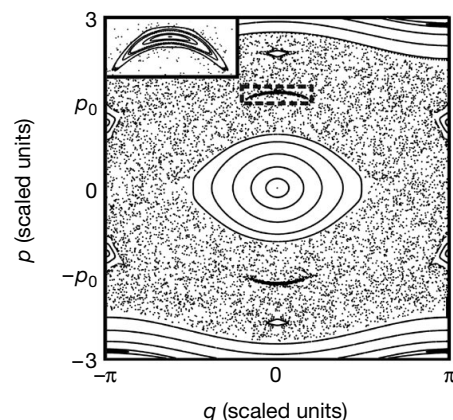


Figure 1 Poincaré section for position p and momentum q of a classical particle in an amplitude-modulated optical lattice. Data are shown for $\kappa = 1.66$ and $\epsilon = 0.29$, showing a mixed phase space. The central region consists of small-amplitude stable motion. Chaos separates this region from two period-one resonances located above and below the centre along $q = 0$. Further out in momentum are two stable regions of motion known as librations, which have a period of twice the modulation period. At the edges are bands of regular motion corresponding to above-barrier motion. The inset is a magnified view of the dashed rectangle.

provide a test of this suggestion in the context of tunnelling in a nonlinear dynamical system.

Atoms in optical potentials^{8–11} provide an ideal test bed to explore quantum nonlinear dynamics. The quantum driven pendulum, a paradigmatic system for the study of ‘quantum chaos’, is realized by a periodically modulated optical standing wave, which can provide a highly non-dissipative, one-dimensional potential. Using ultracold trapped atoms whose action is of the order of Planck’s constant makes quantum effects significant. Atoms in a standing wave, detuned far from an atomic resonance, with sinusoidal amplitude modulation, are described by the centre-of-mass hamiltonian:

$$H(t) = \frac{p^2}{2} + 2\kappa(1 + 2\epsilon \sin\omega t) \sin^2(q/2)$$

where $q = 2kx$ and $p = (2k/m\omega)p_x$ are the scaled position and momentum variables respectively, t is the time, $\kappa = 4U_0^2\omega_R/\omega^2$ is the scaled well depth, ω is the modulation frequency, $\omega_R = \hbar k^2/2m$ is the atomic recoil frequency, U_0 is the well depth in units of $\hbar\omega_R$, and ϵ is the modulation parameter. The position and momentum along the standing wave are x and p_x , m is the mass, $k = 2\pi/\lambda$ is the wavevector, and λ is the wavelength of the optical standing wave. The classical system described by this hamiltonian has a mixed phase space having regions of regular motion embedded in a sea of chaos. The Poincaré section shown in Fig. 1 displays stroboscopically the atomic position and momentum at time intervals equal to the modulation period. The Poincaré section is plotted for experimental parameters $\kappa = 1.66$ and $\epsilon = 0.29$. Islands of regular motion (boomerang-shaped rings—see Fig. 1 inset) surround period-one resonances (fixed points) where the atomic motion remains exactly in phase with the modulation (a period-one resonance rotates 360° around the origin in one modulation period).

Figure 2 shows a graphical representation of these period-one resonances. For small changes in κ and ϵ (within our experimental uncertainties), the period-one regions of regular motion can split into two closely spaced period-two regions. For initial conditions reflecting the Heisenberg uncertainty limit in our experiment, such behaviour would be indistinguishable from period-one motion. The chaotic sea of Fig. 1 (dotted region) is bounded in momentum by the region of regular unbound motion, where atoms have enough energy to move along from one well to the next. The ellipses in Fig. 1 represent Kolmogorov–Arnold–Moser (KAM) surfaces¹², the crossing of which is forbidden by classical mechanics. A detailed experimental and theoretical study of the dynamics of the driven pendulum using cold atoms may be found in previous publications^{10,11}.

The quantum dynamics of a periodically driven hamiltonian system can be described in terms of the eigenstates of the Floquet operator F , which evolves the system in time by one modulation period. The Floquet eigenstates can be associated with regions of regular and irregular motion of the classical map⁴. The localized

Floquet eigenstates do not necessarily overlap well with the regions of regular motion of the classical Poincaré map, except in the semiclassical limit of large action. Initial states localized in the stable region around a fixed point in the Poincaré section can be associated with superpositions of a small number of Floquet eigenstates. For an appropriate choice of parameters the phase space exhibits two period-one fixed points, which for a suitable Poincaré section lie on the momentum axis at $\pm p_0$, as in Fig. 1. For certain values of κ and ϵ there are two dominant Floquet states $|\phi_\pm\rangle$ that are localized on both fixed points but are distinguished by being even or odd eigenstates of the parity operator that changes the sign of momentum. A state localized on just one fixed point is therefore likely to have dominant support on an even or odd superposition of these two Floquet states: $|\psi(\pm p_0)\rangle \approx (|\phi_+\rangle \pm |\phi_-\rangle)/\sqrt{2}$. The evolution is described by repeated application of the Floquet operator. As this is a unitary operator, $F|\phi_\pm\rangle = e^{-i\phi_\pm}|\phi_\pm\rangle$. Thus at a time which is n times the period of modulation, the state initially localized on $+p_0$ evolves to $|\Psi(n)\rangle \approx (e^{-in\phi_+}|\phi_+\rangle + e^{-in\phi_-}|\phi_-\rangle)/\sqrt{2}$. Ignoring an overall phase and defining the separation between Floquet quasi-energies as $\Delta\phi = \phi_- - \phi_+$, we obtain $|\Psi(n)\rangle \approx (|\phi_+\rangle + e^{-in\Delta\phi}|\phi_-\rangle)/\sqrt{2}$. At $n = \pi/\Delta\phi$ periods, the state will form the anti-symmetric superposition of Floquet states and thus is localized on the other fixed point at $-p_0$. In other words, the system has tunneled from one of the fixed points to the other^{4,6}. This is reminiscent of barrier tunnelling between two wells, where a particle in one well, a superposition of symmetric and antisymmetric energy eigenstates, oscillates between wells with a frequency given by the energy difference between the eigenstates.

We prepared atoms in a superposition of Floquet states associated with one of the fixed points as follows. A Bose–Einstein condensate with approximately 3×10^6 sodium atoms was prepared¹³ in a magnetic trap with trapping frequencies $2\omega_x = \sqrt{2}\omega_y = \omega_z = 2\pi \times 33$ Hz, which, using a scattering length of 2.8 nm, results in calculated Thomas–Fermi diameters of 83, 57 and 42 μm , respectively. We release the condensate and adiabatically turn on a one-dimensional standing-wave lattice along z . The gaussian lattice beams are detuned ~ 14 GHz above the atomic D2 resonance ($\lambda = 589$ nm) and have a waist $w \approx 250$ μm , leading to an intensity variation of less than 5% across the atomic cloud. We determine κ from a one-dimensional band structure calculation using the measured oscillation frequency of atoms in the unmodulated standing wave. Its 10% uncertainty is due to uncertainty in measuring the oscillation frequency and to variations in the laser intensity (all reported uncertainties are 1 s.d. combined systematic and statistical uncertainties).

The atoms are loaded into the bottom of the lowest band (quasimomentum equal to 0) of the optical lattice (we have verified that $> 99\%$ of the atomic population is in the lowest band). Each individual well contains a minimum uncertainty wave packet with an r.m.s. spread of 0.6 in scaled units. To load selectively only the region around one resonance, we suddenly shift the position of the standing wave by inducing an appropriate phase shift with an

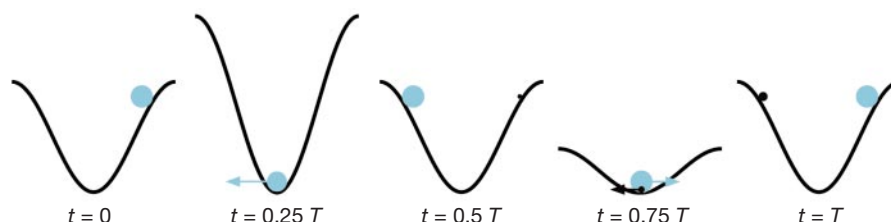


Figure 2 Diagram of period-one resonances of an atom in an amplitude-modulated sinusoidal potential. These resonances correspond to atomic motion that remains in phase with the modulation frequency. The anharmonicity of the sinusoidal potential is compensated by the amplitude modulation for atoms around these resonances, creating

the regions of regular period-one motion shown in Fig. 1. The appearance of the second ball at $0.5T$ depicts tunnelling into the initially empty resonance. T is the modulation period, and t is time.

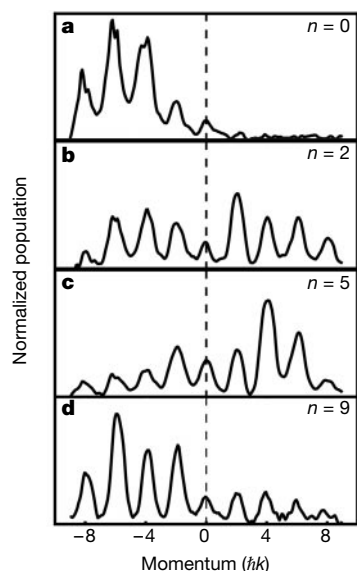


Figure 3 Atomic momentum distributions after n modulation periods, showing dynamical tunnelling. The parameters for this experiment were $\kappa = 1.66$, $\epsilon = 0.29$ and $\omega/2\pi = 250$ kHz.

acousto-optical modulator. Shortly (100–300 ns) before the phase shift, we start (at $t = 0$) to modulate the standing wave at $\omega/2\pi = 220$ –320 kHz. The modulation frequency was chosen to optimize both the overlap between the wave packet and the region of regular motion, and to make the typical action of a particle¹⁰ small enough to observe quantum effects. After the atoms have interacted with the modulated standing wave for a selected number of modulation periods, the standing wave is turned off with the modulation phase chosen so that the resonance lies on the momentum axis at that time. Classically at that phase ($n + 0.25$ or 0.75 periods) the atoms, contained inside the region of regular motion, are at the bottom of the well, moving with maximum momentum, $\pm p_0$ (see Fig. 2). We measure the atomic momentum distribution with absorption imaging after 1.5 ms of free flight. The momentum distribution appears as a set of ‘diffraction’ peaks at integer multiples of $2\hbar k$ due to the atomic coherence over the multiple wells of the optical lattice¹³.

Figure 3 depicts atomic momentum distributions for different interaction times with the modulated standing wave. In Fig. 3a we show the distribution immediately after loading one resonance region. This distribution consists mainly of a pair of diffraction peaks at $4\hbar k$ and $6\hbar k$. Classically, the atoms should remain in the resonance so the stroboscopically measured momentum distribution is unchanged. To the contrary, in Fig. 3b about half of the atoms have appeared with opposite momenta, which correspond to the other resonance region. Most of the atoms are found there by 5.25 modulation periods (Fig. 3c). At 9.25 modulation periods the atoms have returned to the original resonance as shown in Fig. 3d. This transfer of atoms back and forth between the regions of regular motion is coherent dynamical tunnelling.

In Fig. 4a we plot the mean atomic momentum after multiples of the modulation period. The blue (red) data points correspond to turning off the standing wave at the maximum (minimum) of the amplitude modulation (see Fig. 2 at $t = 0.25 T$ and $t = 0.75 T$). We observe an oscillation of the mean momentum indicating the occurrence of dynamical tunnelling for each set of data points. The period-one character of the motion is verified by the reversal of the momentum between the blue and red curves in Fig. 4a, which are separated in time by 0.5 modulation periods. By Fourier-transforming our data, we find the tunnelling period to be

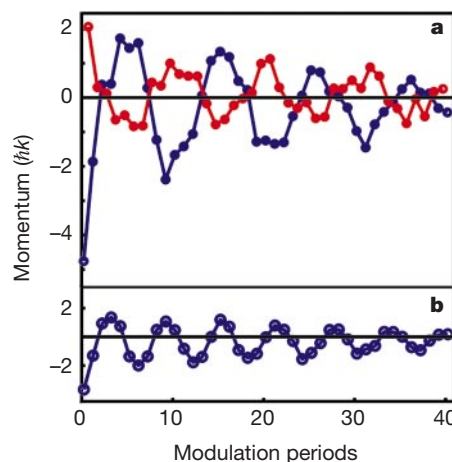


Figure 4 Mean momentum as a function of the number of modulation periods, n . **a**, The two curves correspond to two different ending phases ($n + 0.25$ (blue), $n + 0.75$ (red)) of the modulation (parameters as in Fig. 3). **b**, A slower decay of the tunnelling oscillations is seen for $\epsilon = 0.30$, $\kappa = 1.82$, and $\omega/2\pi = 222$ kHz.

10.3(2) modulation periods where the uncertainty is statistical. We performed a quantum simulation¹¹ for the system parameters of our experiment. The period of the tunnelling oscillation agrees with the experiment within the uncertainty (dominated by the uncertainty in κ); however, the envelope of the oscillations is different. Our simulations assume the presence of only 6 wells, whereas in the experiment about 150 are populated. The initial conditions of the calculations populated a single well (a superposition of all quasimomenta), and showed that the envelope of the oscillations depends sensitively on the number of wells. It is clear that coherence between wells (associated with quasimomentum = 0) must be included to accurately model our experiment. We note that the mean momenta of the blue and the red data points in Fig. 4a are slightly different for the same number of modulation periods. Classical theory predicts slightly different momenta of the fixed points when viewed at the phases corresponding to Fig. 2 at $t = 0.25 T$ and $t = 0.75 T$.

In Fig. 5 we display the atomic momentum distributions corresponding to the blue curve of Fig. 4a, as a function of the number of modulation periods. Because all atoms start on one side of the potential well, at 0.25 modulation periods essentially all atoms have negative momentum. By 1.25 periods the atoms that were loaded into the chaotic region have begun to spread out, forming a broad background, while the other atoms are bound inside the region of regular motion. For subsequent pictures the coherent oscillation between the two regions of regular motion is evident. We note that there is no significant (above-background) zero momentum peak even in the case of approximately zero mean momentum (when the atoms have tunnelled half way). This indicates that at half the tunnel period the system is in a coherent superposition of two distinguishable classical motions: one with positive momenta and one with equal but opposite momenta. We expect this, because quantum Floquet analysis shows that atoms tunnel from one region of regular motion into the other and it is impossible for them to enter the central island of stability at $(p, q) = (0, 0)$.

Classical simulations using a gaussian phase space distribution corresponding to our experimental wave packet (which is always bigger than the classical region of regular motion) show that for the conditions of Fig. 1 and for all experimental conditions, there is no behaviour similar to the observed oscillatory quantum tunnelling. The decay of the tunnelling oscillations may be due to a number of factors. One likely cause for the decay is spatial variations of κ , which lead to a dephasing of the tunnelling oscillations. Another possibil-

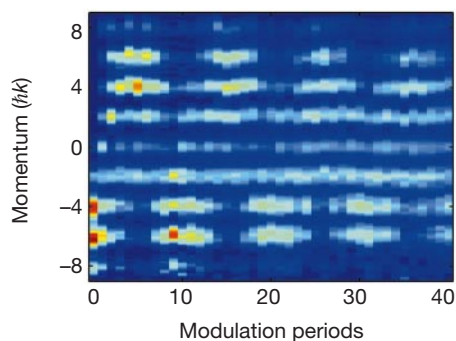


Figure 5 Momentum distributions as a function of the number of modulation periods, showing the tunnelling oscillation between negative and positive momenta. We note that the zero-momentum state remains mostly unpopulated, even when the mean momentum is zero. The colour coding ranges from blue to red for atomic populations ranging from small to large.

ity is that the tails of the oscillating quantum wave packets may extend outside the region of regular motion, allowing the atoms to 'leak out' into the classically chaotic region. This possibility is currently under investigation. The contribution of multiple Floquet states could lead to complicated multi-frequency oscillations, and an envelope for the tunnel oscillations appearing as decay, as observed for some parameters in our simulations. The effects of spontaneous emission and atom-atom interactions should be small.

Quantum theory predicts dynamical tunnelling to occur for various values of the scaled well depth κ , the modulation parameter ϵ and modulation frequency ω , and also predicts a strong sensitivity of the tunnelling period and amplitude on these parameters. For $\epsilon = 0.23$, $\kappa = 1.75$ and $\omega/2\pi = 250$ kHz we measured a tunnelling period of approximately 13 modulation periods. As shown in Fig. 4b, for $\epsilon = 0.30$, $\kappa = 1.82$ and $\omega/2\pi = 222$ kHz we find a tunnelling period of 6 modulation periods with a significantly longer decay time than in Fig. 4a. We have also experimentally observed an increase in the tunnelling period when κ is decreased and when all other parameters are held constant. This is the opposite of what one would expect for spatial barrier tunnelling.

Our observation of dynamical tunnelling of atoms in a modulated standing wave opens the door to further studies in quantum nonlinear dynamics. By varying the hamiltonian parameters and the initial conditions, we observe dynamical tunnelling for a variety of mixed phase space configurations. This may be 'chaos-assisted tunnelling', and a tunnelling rate that varies wildly as system parameters are changed would be a signature of such tunnelling¹. By introducing noise or spontaneous emission in a controlled manner, we could systematically investigate the role of decoherence in tunnelling and explore the classical limit of chaotic systems. By carefully following the evolution of wave packets loaded into the chaotic region from a Bose-Einstein condensate, we could probe 'quantum chaos' with the unprecedented resolution afforded by minimum uncertainty wave packets.

During the preparation of this Letter, we learned of an experiment¹⁴ reporting tunnelling of a different motional state in a similar system. □

Received 10 May; accepted 12 June 2001.

1. Tomsovic, S. Tunneling and chaos. *Physica Scripta T* **90**, 162–165 (2001).
2. Caldeira, A. O. & Leggett, A. J. Quantum tunneling in a dissipative system. *Ann. Phys.* **149**, 374–456 (1983).
3. Davis, M. J. & Heller, E. J. Quantum dynamical tunneling in bound states. *J. Chem Phys.* **75**, 246–254 (1981).
4. Dyrting, S., Milburn, G. J. & Holmes, C. A. Nonlinear quantum dynamics at a classical second order resonance. *Phys. Rev. E* **48**, 969–978 (1993).

5. Haake, F., Kus, M. & Scharf, R. Classical and quantum chaos for a kicked top. *Z. Phys. B* **65**, 381–395 (1987).
6. Sanders, B. C. & Milburn, G. J. The effect of measurement on the quantum features of a chaotic system. *Z. Phys. B* **77**, 497–510 (1989).
7. Habib, S., Shizume, K. & Zurek, W. H. Decoherence, chaos, and the correspondence principle. *Phys. Rev. Lett.* **80**, 4361–4365 (1998).
8. Graham, R., Schlautmann, M. & Zoller, P. Dynamical localization of atomic-beam deflection by a modulated standing light wave. *Phys. Rev. A* **45**, R19–R22 (1992).
9. Moore, F. L., Robinson, J. C., Bharucha, C. F., Sundaram, B. & Raizen, M. G. Atom optics realization of the quantum δ -kicked rotor. *Phys. Rev. Lett.* **75**, 4598–4601 (1995).
10. Hensinger, W. K., Truscott, A. G., Upcroft, B., Heckenberg, N. R. & Rubinsztein-Dunlop, H. Atoms in an amplitude-modulated standing wave—dynamics and pathways to quantum chaos. *J. Opt. B* **2**, 659–667 (2000).
11. Hensinger, W. K. *et al.* Experimental study of the quantum driven pendulum and its classical analog in atom optics. *Phys. Rev. A* (in the press).
12. Arnold, V. I. *Mathematical Methods of Classical Mechanics* (Springer, New York, 1979).
13. Kozuma, M. *et al.* Coherent splitting of Bose-Einstein condensed atoms with optically induced Bragg diffraction. *Phys. Rev. Lett.* **82**, 871–875 (1999).
14. Steck, D. A., Oskay, W. H. & Raizen, M. G. Observation of chaos-assisted tunneling between islands of stability. *Science* (in the press).

Acknowledgements

We thank C. Holmes for discussions. The NIST group was supported by the ONR, NASA and ARDA, and the University of Queensland group was supported by the ARC. A.B. was partially supported by DGA (France), and H. H. was partially supported by the A. v. Humboldt Foundation. W.K.H. and B.U. thank NIST for hospitality during the experiments.

Correspondence and requests for materials should be addressed to W.K.H. (e-mail: hensinger@physics.uq.edu.au).

Quantum interference of superfluid ³He

R. W. Simmonds, A. Marchenkov, E. Hoskinson, J. C. Davis & R. E. Packard

Physics Department, University of California, Berkeley, California 94720, USA

Celebrated interference experiments have demonstrated the wave nature of light¹ and electrons², quantum interference being the manifestation of wave-particle duality. More recently, double-path interference experiments have also demonstrated the quantum-wave nature of beams of neutrons³, atoms⁴ and Bose-Einstein condensates⁵. In condensed matter systems, double-path quantum interference is observed in the d.c. superconducting quantum interference device⁶ (d.c. SQUID). Here we report a double-path quantum interference experiment involving a liquid: superfluid ³He. Using a geometry analogous to the superconducting d.c. SQUID, we control a quantum phase shift by using the rotation of the Earth, and find the classic interference pattern with periodicity determined by the ³He quantum of circulation.

Our basic interferometer topology is shown in Fig. 1a. Schematically, the device is a circular loop of radius R which includes two superfluid ³He Josephson weak links^{7,8}. These weak links each consist of a 65×65 array of 100-nm apertures etched in a 60-nm-thick silicon nitride membrane. Similar arrays have previously been shown to be characterized by a current-phase relation given⁹ by the Josephson formula:

$$I = I_c \sin \phi \quad (1)$$

Here I is the mass current flowing through the array, ϕ is the quantum phase difference across the array, and I_c is the critical current characterizing the array.

The interferometer is predicted to behave as a single weak link with an effective critical current, I_c^* (see Methods). If the inter-

ferometer is rotated at angular velocity Ω , I_c^* is modulated by an interference term¹⁰

$$I_c^* = 2I_c \left| \cos \left(\pi \frac{2\Omega \cdot \mathbf{A}}{\kappa_3} \right) \right| \quad (2)$$

where $\mathbf{A}$ is the area vector of the loop and $\kappa_3 \equiv h/(2m_3)$ is the ^3He quantum of circulation. Thus the effective critical current is predicted to be modulated with a period determined by the rotation flux through the interferometer, analogous to the manner in which magnetic flux modulates the critical current of a d.c. SQUID.

The interference occurs because the phase of the superfluid wavefunction receives equal and opposite shifts in the two arms of the interferometer. The phase shift arising from the rotation flux is identical in form to that in other neutral-matter quantum rotation interferometers^{3,11}, as well as in the optical Sagnac interferometer¹² if the photon effective mass is taken to be $\hbar\omega/c^2$.

The goal of our experiment is to demonstrate two ideas. (1) That the superfluid interferometer is indeed characterized as a single weak link with a well defined critical current I_c^* , and (2) that by changing the orientation of the plane of the loop with respect to the Earth's rotation vector, the current I_c^* can be modulated according to the interference term in equation (2). I_c^* should exhibit periodicity of $(2\Omega \cdot \mathbf{A})/\kappa_3$.

A geometrically more accurate sketch of our interferometer is shown in Fig. 1b. The quantum phase difference across the

interferometer should evolve according to the equation¹³:

$$\Phi = - \int \frac{2m_3}{\rho\hbar} P(t) dt \quad (3)$$

where ρ is the density of the liquid and P is the pressure differential across the ends of the device. So if the interferometer behaves as a single Josephson weak link, a static pressure P_0 applied across the interferometer will cause the quantum phase to increase linearly in time, leading to mass current oscillations at a Josephson frequency⁸:

$$\omega_j = \frac{2m_3}{\hbar\rho} P_0 \quad (4)$$

As described in Fig. 1 legend, the deflection of the flexible membrane reveals both the pressure across the interferometer and the mass current through it. We have developed a feedback method that permits us to drive the system at constant pressure by applying a time-varying voltage to the membrane. We select a pressure for which the Josephson frequency lies near 270 Hz, a spectral region away from parasitic acoustic noise lines in the displacement transducer.

Figure 2 shows a Fourier transform of the mass current through the interferometer that results from the constant-pressure drive being applied for about 6 seconds. A sharp peak at 273 Hz is clearly visible. This corresponds to the Josephson oscillation. This

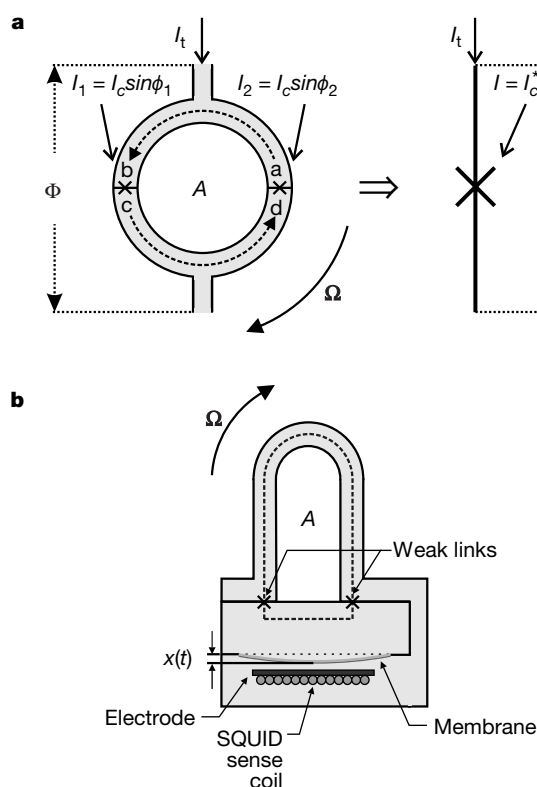


Figure 1 Two views of our superfluid quantum interferometer. **a**, Schematic diagram of the basic interferometer loop. As shown in the Methods section, the total current through the two sides of the loop (I) is predicted to depend on the total phase difference across the loop, Φ , as given by equation (1) with I_c replaced by I_c^* (given by equation (2)). **b**, The interferometer geometry used in this experiment. Although the two weak links are moved closer together, the topology is unchanged and I_c^* is still given by equation (2). The nominal 'loop' area, A , is $\sim 6 \text{ cm}^2$, and the tube cross-sectional diameter is $\sim 0.3 \text{ cm}$. Pressure differences can be applied to the flexible metallized membrane (and hence across the two weak links) by applying voltages between it and the adjacent rigid electrode. The membrane is coated with a superconducting film. The sense coil contains a

trapped persistent current which is partially shunted into the input coil of a d.c. SQUID if the membrane moves²⁶. Knowledge of the deflection of the membrane, $x(t)$, reveals the pressure head across the interferometer. The rate of change of the deflection reveals the mass current through the interferometer. The noise in the displacement sensor is $2 \times 10^{-15} \text{ m Hz}^{-1/2}$. We use a feedback system that applies a voltage to the electrode such that the pressure across the interferometer is fixed at some predetermined value, typically near 1.5 mPa. Temperature is monitored with two thermometers, one using platinum nuclear magnetic resonance and a second using lanthanum cerium magnesium nitrate (LCMN). The experiment is performed at zero ambient pressure.

frequency agrees with equation (4) to within the systematic uncertainty ($\sim 10\%$) in the electrostatically derived calibration of the pressure scale. We have determined that the oscillation frequency scales linearly with P_0 at least up to 1 kHz, and that there is no higher-harmonic signal detectable with the present signal-to-noise ratio. This implies that the overall current–phase relation is sine-like, and that the two separated arrays are phase coherent. The area under the Josephson oscillation peak in Fig. 2 is proportional to I_c^* , the critical current for the interferometer.

This oscillation signal results from the quantum coherence among all the apertures in each array, as well as between both arrays. Thus, not only are both weak links themselves coherent, but the $\sim 10^{22}$ atoms within the torus are also quantum phase coherent with them. This demonstrates the first point, that the interferometer behaves as a single Josephson weak link with a well defined critical current, I_c^* .

Furthermore, weak-link arrays have previously been shown¹⁴ to exhibit a characteristic rigid-pendulum oscillation mode whose small-amplitude frequency ω_p is proportional to $\sqrt{I_c}$. We have observed this normal mode motion in our interferometer and find $\omega_p \propto \sqrt{I_c^*}$, again confirming that the device behaves as a single weak link characterized by equations (1) and (3).

The second goal of our experiment is to see if phase gradients, generated by absolute rotation of the loop, create an interference pattern to modulate I_c^* , as predicted in equation (2). We chose the loop area A to be 6 cm^2 so that the rotation of the Earth, Ω_E , can itself lead to several periods of the modulation pattern. The plane of the loop is vertical—that is, the vector $\mathbf{A}$ points horizontally in the laboratory. Thus we can vary the rotation flux in the loop, $\Omega_E \cdot \mathbf{A}$, by reorienting the interferometer about a vertical axis in the laboratory frame. This reorientation technique was used previously in superfluid ^4He to show that rotation flux could vary the phase-slip critical velocity in an aperture^{15,16}. Other experiments employed

reorientation of a superfluid ^3He loop to change the frequency in a nonlinear Helmholtz oscillator^{17,18} containing a single Josephson weak link.

Figure 3 shows the result of our reorientation experiment. We plot the normalized measured current magnitude I_c^* as a function of $(2\Omega \cdot \mathbf{A})/\kappa_3$. The data reveal the double-path quantum interference pattern. The solid curve is the interference pattern expected from equation (2). The periodicity of the observed interference pattern is found to be as predicted: $(2\Omega \cdot \mathbf{A})/\kappa_3$, within the experimental precision (which is limited by knowledge of the loop area).

The interference pattern predicted by equation (2) and shown in Fig. 3 is the central feature of this experiment. The device displays a remarkable phenomenon: two-path quantum interference in a liquid. The pattern shows that the interferometer is indeed the superfluid equivalent of a d.c. SQUID.

It is natural to ask if the superfluid quantum interference device could be developed into a sensitive rotation sensor^{19–21}, perhaps to perform meaningful geodesy measurements or experiments on general relativity. Following the analysis²² of superconducting d.c. SQUIDS, the intrinsic noise in this type of device arises from Nyquist noise in the various dissipative processes associated with the weak links themselves²³. The potential intrinsic sensitivity can only be reached if other extrinsic noise sources—such as temperature drifts, environmental noise and electronic SQUID readout noise—are reduced by orders of magnitude from the values in the present experiment.

We have demonstrated a double-path superfluid interferometer which is the analogue of a d.c. SQUID. The entire interferometer exhibits dynamic behaviour (Josephson oscillations and pendulum motion) similar to that of a single weak link. Owing to quantum interference, the effective critical current is modulated by rotation flux through the enclosed area. Our double-path interference

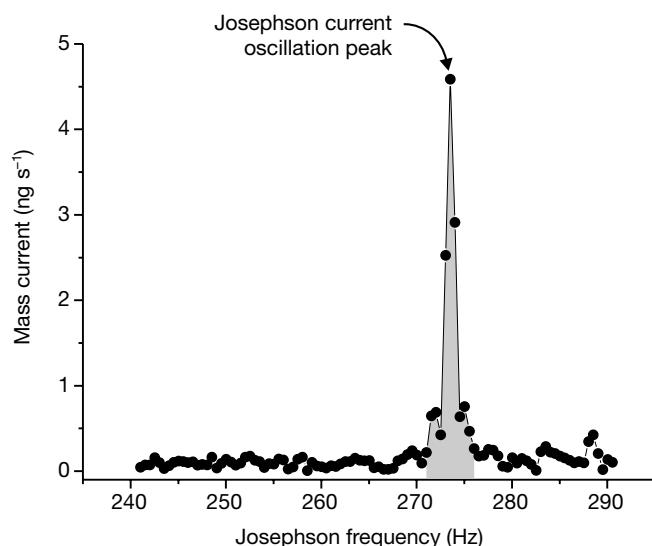


Figure 2 The spectrum of the mass current during a 2-s interval of the data stream from the SQUID position transducer. A typical data stream is 6 s long, limited by the magnitude of the voltage we can apply to the diaphragm. Due to imperfect pressure regulation, the Josephson peak drifts slightly during the 6-s data stream. So we break the transient into 2-s intervals, and use a fast Fourier transform (FFT) routine to produce a figure like that above for each interval. The peak is due to Josephson oscillations at 273 Hz. We compute a number proportional to the area under the peak (shaded in the figure). At a fixed orientation we record multiple data streams (typically 17), and average the resultant areas. This average is taken as the measure of I_c^* , the critical current of the interferometer.

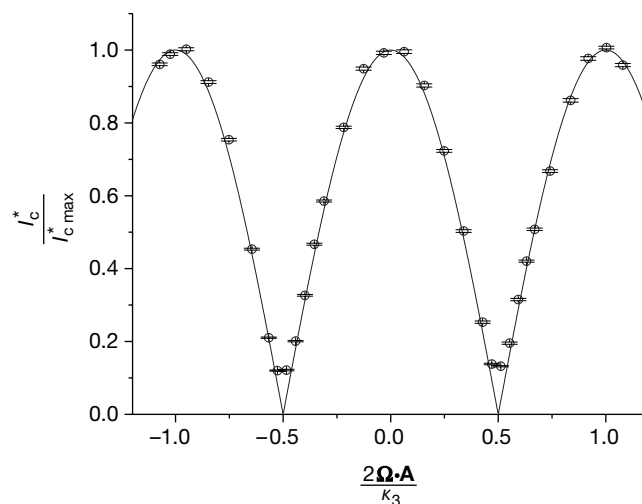


Figure 3 The interference pattern of a superfluid quantum interference gyroscope. The figure shows a plot of the effective critical current I_c^* as a function of $(2\Omega \cdot \mathbf{A})/\kappa_3$, where κ_3 is the quantum of circulation (see equation (2)). The error bars, which are the standard error of typically 100 FFTs, are smaller than the size of the plotted points. The temperature was approximately $0.8T_c$, where T_c is the critical temperature of the superfluid. We have applied a temperature correction (no greater than 4%) to account for systematic errors from small temperature drifts recorded by the LCMN during the course of the measurement. The rotation flux was varied by reorienting the normal to the loop's plane through an angle of $\pm \pi/2$ with respect to the east–west direction. Then the rotation flux $\Omega \cdot \mathbf{A}$ is equal to $\Omega_E A \cos \theta_L \sin \Theta$, where Θ is the direction of $\mathbf{A}$ with respect to an east–west line, and $\theta_L \approx 38^\circ$ is the latitude of Berkeley. The solid line is a plot of equation (2).

experiment advances the close analogy between the macroscopic quantum physics of superconductivity and superfluidity. Both systems exhibit persistent currents, quantized circulation (fluid and magnetic), $\sin\phi$ weak links and now, double-path quantum interference. □

Methods

To derive equation (2) we follow a heuristic method similar to that which Feynman²⁴ applied to the analogous superconducting case. Consider the interferometer in Fig. 1a to be rotating at angular velocity Ω . The total external current passing through the interferometer may be written as:

$$I_t = I_c \sin\phi_1 + I_c \sin\phi_2 = 2I_c \cos\left(\frac{\phi_1 - \phi_2}{2}\right) \sin\left(\frac{\phi_1 + \phi_2}{2}\right) \quad (5)$$

We assume that the fluid in the torus is characterized at every point by a quantum phase factor which is macroscopically coherent. The rotation of the torus entrains the superflow which is then almost that of a solid body, $v_s = \Omega R$. Around a closed path in the interferometer, as indicated by the dotted line in Fig. 1, the phase can change only in multiples of 2π .

$$\oint \nabla\phi \cdot d\mathbf{l} = 2\pi n = \int_a^b \frac{2m_3}{\hbar} \mathbf{v}_s \cdot d\mathbf{l} + \phi_1 + \int_c^d \frac{2m_3}{\hbar} \mathbf{v}_s \cdot d\mathbf{l} - \phi_2 = -\frac{2m_3}{\hbar} 2\Omega \pi R^2 + \phi_1 - \phi_2 \quad (6)$$

The limit points, $a-d$, in the integrals are shown in Fig. 1a. In equation (6) we have used the concept²⁵ that the phase gradient is proportional to superfluid velocity, $\nabla\phi = (2m_s \mathbf{v}_s)/\hbar$. Equation (6) may be solved for $\phi_1 - \phi_2$, which, when inserted into equation (5) gives

$$I_t = I_c^* \sin\Phi \quad (7)$$

where $\Phi \equiv (\phi_1 + \phi_2)/2$ is the phase difference across the ends of the interferometer. We see that the rotating interferometer behaves as a single Josephson weak link, whose maximum current is given by the quantum interference term:

$$I_c^* = 2I_c \left| \cos\left(\pi \frac{2\Omega \cdot \mathbf{A}}{\kappa_3}\right) \right| \quad (8)$$

Here $\kappa_3 \equiv \hbar/(2m_3)$ is the quantum of circulation, $\mathbf{A}$ is the interferometer area vector, normal to the plane, and we have included a scalar product describing the more general case when the rotation axis is not parallel to $\mathbf{A}$.

Received 5 March; accepted 14 May 2001.

- Young, T. A *Course of Lectures in Natural Philosophy and the Mechanical Arts* Vol. 1, 364 (London, 1845); also as facsimile edition (New York, 1971).
- Davison, C. J. Are Electrons Waves? *Franklin Inst. J.* **205**, 597 (1928).
- Werner, S. A., Studenmann, J. L. & Colella, R. Effect of Earth's rotation on the quantum mechanical phase of the neutron. *Phys. Rev. Lett.* **42**, 1103–1106 (1979).
- Keith, D. W., Ekstrom, C. R., Turchette, Q. A. & Pritchard, D. E. An interferometer for atoms. *Phys. Rev. Lett.* **66**, 2693–2696 (1991).
- Andrews, M. R. *et al.* Observation of interference between two Bose condensates. *Science* **275**, 637–641 (1997).
- Barone, A. & Paterno, G. *Physics and Application of the Josephson Effect* (Wiley, New York, 1982).
- Varoquaux, E. & Avenel, O. Josephson effect and quantum phase slippage in superfluids. *Phys. Rev. Lett.* **60**, 416–419 (1988).
- Perverse, S. V., Loshak, A., Backhaus, S., Davis, J. C. & Packard, R. E. Quantum oscillations between two weakly coupled reservoirs of superfluid ^3He . *Nature* **388**, 449–451 (1997).
- Marchenkov, A. *et al.* Bi-state superfluid ^3He weak links and the stability of Josephson π states. *Phys. Rev. Lett.* **83**, 3860–3863 (1999).
- Packard, R. E. & Vitale, S. Principle of superfluid-helium gyroscopes. *Phys. Rev. B* **46**, 3540–3549 (1992).
- Gustavson, T. L., Bouyer, P. & Kasevitch, M. A. Physical rotation measurements with an atom interferometer gyroscope. *Phys. Rev. Lett.* **78**, 2046–2049 (1997).
- Stedman, G. E. Ring-laser tests of fundamental physics and geophysics. *Rep. Prog. Phys.* **60**, 615–687 (1997).
- Anderson, P. W. in *Progress in Low Temperature Physics* (ed. Gorter, C. J.) 1–44 (North Holland, Amsterdam, 1967).
- Marchenkov, A., Simmonds, R. W., Davis, J. C. & Packard, R. E. Observation of the Josephson plasma mode for a superfluid ^3He weak link. *Phys. Rev. B* **61**, 4196–4199 (2000).
- Schwab, K., Bruckner, N. & Packard, R. E. Detection of the Earth's rotations using superfluid phase coherence. *Nature* **386**, 585–587 (1997).
- Avenel, O., Hakonen, P. & Varoquaux, E. Detection of the rotation of the Earth with a superfluid gyrometer. *Phys. Rev. Lett.* **78**, 3602–3605 (1997).
- Mukharsky, Yu., Varoquaux, E. & Avenel, O. Current-phase relationship measurements in the flow of superfluid ^3He through a single orifice. *Physica B* **280**, 130–131 (2000).
- Mukharsky, Yu., Avenel, O. & Varoquaux, E. Rotation measurements with a superfluid ^3He gyrometer. *Physica B* **284–288**, 287–288 (2000).
- Rowe, C. H. *et al.* Design and operation of a very large ring laser gyroscope. *Appl. Opt.* **38**, 2516–2523 (1999).
- Gustavson, T. L., Landragin, A. & Kasevich, M. A. Rotation sensing with a dual atom interferometer Sagnac gyroscope. *Class. Quant. Gravity* **17**, 2385–2398 (2000).

- Herring, T. A. The rotation of the Earth. *Rev. Geophys. Suppl.* **29**, 172–175 (1991).
- Clarke, J. in *SQUID Sensors: Fundamentals, Fabrication and Applications* (ed. Weinstock, H.) (Kluwer Academic, 1996).
- Simmonds, R. W., Marchenkov, A., Vitale, S., Davis, J. C. & Packard, R. E. New flow dissipation mechanisms in superfluid ^3He . *Phys. Rev. Lett.* **84**, 6062–6065 (2000).
- Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures in Physics* Vol. 3, Ch. 21 (Addison Wesley, Reading, Massachusetts, 1963).
- Tilley, D. R. & Tilley, J. *Superfluidity and Superconductivity* 3rd edn 171 (Hilger, Bristol, 1990).
- Paik, H. J. Superconducting tunable-diaphragm transducer for sensitive acceleration measurements. *J. Appl. Phys.* **47**, 1168–1178 (1976).

Acknowledgements

We thank S. Vitale and K. Penanen for discussions; Y. Sato for assistance; A. Loshak for making the aperture arrays; and E. Crump, D. Mathews and C. Ku for assistance in improving noise conditions in our building. This work was supported in part by NASA, the Office of Naval Research, the National Science Foundation, and the Miller Institute for Basic Research (J.C.D.).

Correspondence and requests for materials should be addressed to R.P. (e-mail: packard@socrates.berkeley.edu).

Coexistence of superconductivity and ferromagnetism in the *d*-band metal ZrZn_2

C. Pfleiderer*, M. Uhlarz*, S. M. Hayden†, R. Vollmer*, H. v. Löhneysen*‡, N. R. Bernhoeft§ & G. G. Lonzarich||

* *Physikalisches Institut, Universität Karlsruhe, D-76128 Karlsruhe, Germany*

† *H. H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK*

‡ *Forschungszentrum Karlsruhe, Institut für Festkörperphysik,*

D-76021 Karlsruhe, Germany

§ *DRFMC-SPSMS, CEA Grenoble, F-38054 Grenoble, Cedex 9, France*

|| *Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK*

It has generally been believed that, within the context of the Bardeen–Cooper–Schrieffer (BCS) theory of superconductivity, the conduction electrons in a metal cannot be both ferromagnetically ordered and superconducting^{1,2}. Even when the superconductivity has been interpreted as arising from magnetic mediation of the paired electrons, it was thought that the superconducting state occurs in the paramagnetic phase^{3,4}. Here we report the observation of superconductivity in the ferromagnetically ordered phase of the *d*-electron compound ZrZn_2 . The specific heat anomaly associated with the superconducting transition in this material appears to be absent, and the superconducting state is very sensitive to defects, occurring only in very pure samples. Under hydrostatic pressure superconductivity and ferromagnetism disappear at the same pressure, so the ferromagnetic state appears to be a prerequisite for superconductivity. When combined with the recent observation of superconductivity in UGe_2 (ref. 4), our results suggest that metallic ferromagnets may universally become superconducting when the magnetization is small.

The compound ZrZn_2 was first investigated by Matthias and Bozorth⁵ in the 1950s, who discovered that it was ferromagnetic despite being made from nonmagnetic, superconducting constituents. ZrZn_2 crystallizes in the C15 cubic Laves structure, as shown in the inset of Fig. 1, with lattice constant $a = 7.393 \text{ \AA}$. The Zr atoms form a tetrahedrally coordinated diamond structure and the magnetic properties of the compound derive from the Zr *4d* orbitals, which have a significant direct overlap⁶. Ferromagnetism develops

below the Curie temperature $T_{\text{FM}} = 28.5 \text{ K}$ with a small ordered moment $\mu_s = 0.17 \mu_B$ per formula unit. ZrZn_2 has a large electronic heat capacity at low temperatures $C/T = 47 \text{ mJ K}^{-2} \text{ mol}^{-1}$ signalling an abundance of low-energy magnetic excitations in addition to conventional spin-wave contributions⁷. The low Curie temperature and small ordered moment render ZrZn_2 special among stoichiometric ferromagnetic metals and indicate that the compound is close to a ferromagnetic quantum critical point. The most remarkable magnetic property of ZrZn_2 is the effect of a magnetic field on the ordered moment⁵. Figure 1 shows the magnetization of ZrZn_2 as a function of magnetic field. At $T = 1.75 \text{ K}$ a relatively small field $\mu_0 H = 0.05 \text{ T}$ is required to form a single ferromagnetic domain. On further increasing the field, the ordered moment is rapidly increased with a field of 6 T causing a 50% increase in the ordered moment, which is unsaturated up to 35 T, the highest field measured⁸. This behaviour contrasts strongly with the elemental ferromagnets Fe, Ni and Co, in which, after a single domain is formed, fields applied parallel to the easy axis have only a small effect on the ordered moment.

The proximity of ZrZn_2 to a ferromagnetic quantum critical point has led to numerous proposals that it might be a superconductor^{9–11}. We have therefore studied the resistivity, a.c. (differential) susceptibility and specific heat capacity of high-purity ZrZn_2 at low temperatures. Figure 2 shows measurements of the basic physical properties of ZrZn_2 for two samples of differing quality. The highest-quality sample (C) has a low-temperature residual resistivity of $0.62 \mu\Omega \text{ cm}$ consistent with charge carrier mean free paths of several hundred ångströms. Figure 2a shows that there is a rapid drop in the electrical resistivity below a temperature of $T_{\text{SC}} = 0.29 \text{ K}$, suggesting a transition to superconductivity. The resistance drop is absent in the lower-quality sample (sample A), which has a residual impurity scattering rate five times higher than sample C. The application of a field of 0.2 T suppresses the drop, as expected for a superconducting transition. A second signature of superconductivity is the diamagnetic screening of a magnetic field, which can be observed in the a.c. susceptibility. Because ZrZn_2 is

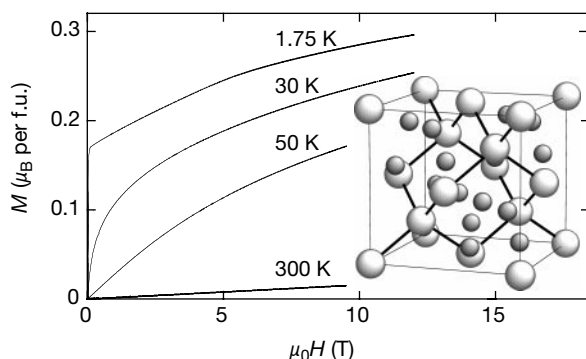


Figure 1 Magnetization curves and crystal structure of ZrZn_2 . A spontaneous moment μ_s develops below the Curie temperature $T_{\text{FM}} = 28.5 \text{ K}$. For $T = 1.75 \text{ K}$, we find a spontaneous moment $\mu_s = 0.17 \mu_B$ per formula unit (f.u.) and a coercive field (determined from a hysteresis loop) of $\mu_0 H^* = 2 \times 10^{-3} \text{ T}$ (μ_B : Bohr magneton). The increase of the ordered moment with field demonstrates the unsaturated nature of the ordered moment and indicates the presence of a large longitudinal susceptibility. Inset, the cubic C15 Laves phase unit cell of ZrZn_2 , where the Zr atoms (large spheres) form a tetrahedrally coordinated diamond structure. The samples investigated here were grown by a directional-cooling method²³ in which zone-refined Zr and Zn were melted in a high-purity Y_2O_3 crucible sealed inside a Ta pressure container. Single crystals were cut from the resulting ingot by spark erosion. Neutron and Laue X-ray diffraction spectra were consistent with the C15 structure previously reported. Electron-probe microanalysis revealed no impurity phases or segregations of the constituent elements at the resolution limit of 0.5%.

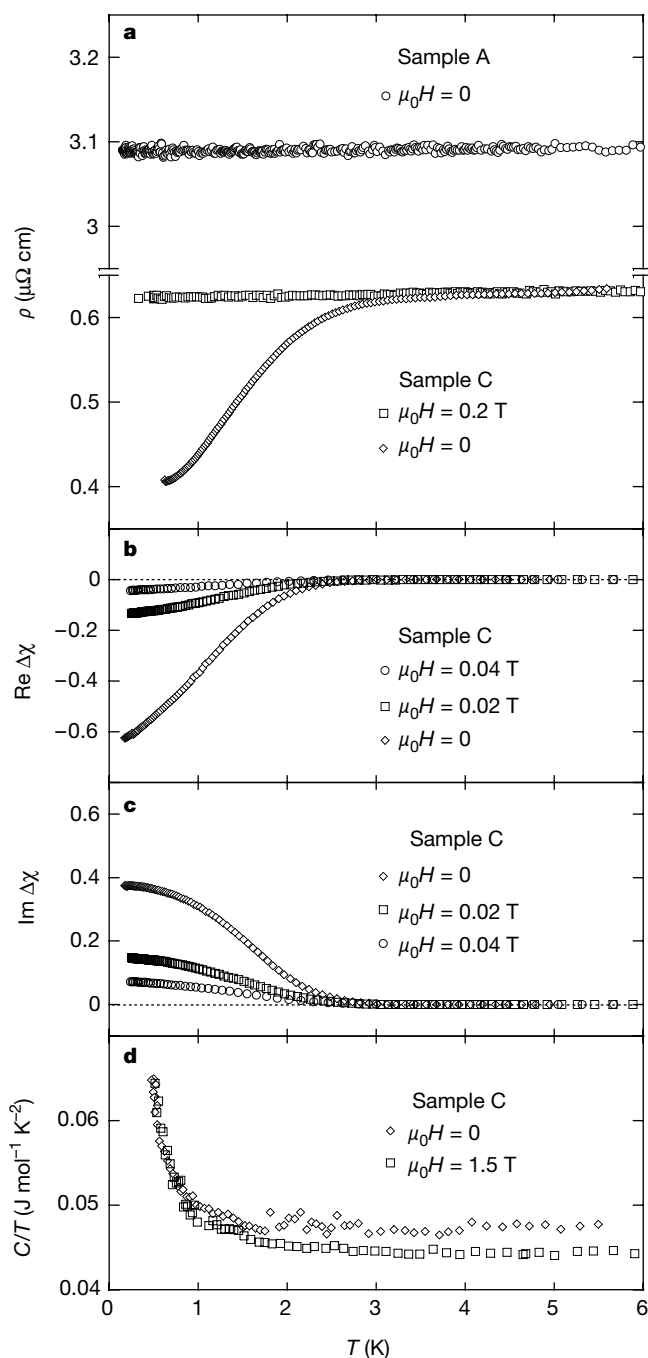


Figure 2 Evidence of superconductivity in high-purity single crystals of weakly ferromagnetic ZrZn_2 . **a**, Resistivity versus temperature, $\rho(T)$, at $\mu_0 H = 0$ and 0.2 T, respectively, as measured in a dilution refrigerator using a four-terminal a.c. technique. Sample A (non-superconducting) is of lower quality than sample C, that is, sample A has a five times higher residual resistivity ρ_0 . **b**, Real (reactive) part of the low-amplitude a.c. susceptibility in SI units versus temperature for an a.c. amplitude of 10^{-6} T , $\text{Re} \Delta \chi(T)$, at various superposed d.c. fields after subtraction of the background due to the ferromagnetism. Measurements were performed at low frequencies with a miniature susceptometer comprised of a primary and a balanced pair of secondary coils. The background was extrapolated from the T -independent susceptibility above T_{SC} , the superconducting ordering temperature. For the lowest excitation amplitudes $\text{Re} \Delta \chi(T)$ approaches almost complete diamagnetic shielding: -0.65 (near -1). An analogous diamagnetic contribution is absent in the low-quality sample (sample A) (not shown). **c**, Imaginary (dissipative) part of the a.c. susceptibility, $\text{Im} \Delta \chi(T)$. Below T_{SC} a substantial contribution develops, typical of type II superconductors. **d**, Specific heat divided by temperature versus temperature, $C(T)/T$, at $\mu_0 H = 0$ and 1.5 T as measured in a dilution refrigerator by a conventional heat pulse relaxation technique. The field-independent upturn below 0.15 K is related to nuclear quadrupolar contributions of Zr.

ferromagnetic, the a.c. susceptibility, $\chi = dM/dH$, has a large component that is due to ferromagnetic domain alignment at low fields. We have therefore subtracted a temperature-independent background, measured above T_{SC} , from our susceptibility data. Figure 2 shows the resulting real (reactive) and imaginary (dissipative) parts of the susceptibility, $\text{Re}\Delta\chi$ and $\text{Im}\Delta\chi$ respectively. A strong diamagnetic signal in $\text{Re}\Delta\chi$ associated with superconducting screening is observed below $T_{SC} = 0.29$ K. For the lowest excitation amplitudes $\text{Re}\Delta\chi$ approaches -0.65 as $T \rightarrow 0$, comparable with the ideal value of -1 . A concomitant increase in the dissipative component $\text{Im}\Delta\chi$ is observed, as for other type II superconductors. The a.c. susceptibility of the low-quality sample (sample A) does not exhibit any signs of a diamagnetic contribution. We have also performed SQUID (superconducting quantum interference device) d.c. magnetization measurements below 1 K on sample C (not shown). The zero-field-cooled d.c. magnetization corresponds quantitatively with $\text{Re}\Delta\chi$, as expected, while the field-cooled magnetization shows a negligible Meissner effect (flux expulsion), as do oxide superconductors. The low-temperature specific heat is shown in Fig. 2d. In the temperature range around T_{SC} no anomaly is observed, and the normal-state electronic contribution of $C/T = 47 \text{ mJ mol}^{-1} \text{ K}^{-2}$ prevails. The specific heat shows a pronounced increase below 0.15 K, which is unaffected by the application of a 1.5-T magnetic field, suggesting that it is not associated with the superconductivity, but rather with a nuclear quadrupolar contribution.

The field dependence of the a.c. susceptibility allows us to identify the critical field $\mu_0 H_{c2}$ below which superconductivity appears. A typical trace showing the rapid increase of $\text{Im}\chi$ as the field is lowered through $\mu_0 H_{c2}$ is shown in the inset of Fig. 3. From the temperature dependence of this onset field we have determined the superconducting phase diagram shown in Fig. 3. At the lowest temperatures investigated, $T = 25$ mK, we find $\mu_0 H_{c2} = 0.4$ T, which corresponds to a superconducting coherence length of $\xi = 290$ Å.

It has been known for some time that the ferromagnetism in ZrZn_2 is rapidly suppressed under pressure¹². This, and the prediction that superconductivity is controlled by the quantum critical point where ferromagnetism disappears at zero temperature, led us

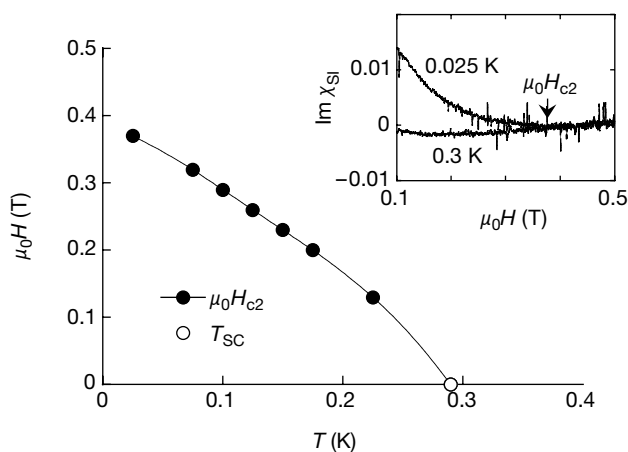


Figure 3 Superconducting phase diagram of ZrZn_2 . Temperature dependence of the upper critical magnetic field $\mu_0 H_{c2}$ as derived from the dissipative part of the a.c. susceptibility, $\text{Im}\chi$. The value of $\mu_0 H_{c2}(T = 0)$ corresponds to a coherence length of $\xi = 290$ Å. We note that $\mu_0 H_{c2}$ is larger than the values of elemental Zr and Zn by two orders of magnitude, where T_{SC} of Zr and Zn are 0.875 K and 0.65 K, respectively. Inset, typical field dependence of the dissipative part of the a.c. susceptibility, $\text{Im}\chi$, used to determine $\mu_0 H_{c2}$.

to perform high-pressure studies. Figure 4 summarizes the effect of pressure on the Curie temperature, T_{FM} , and the superconducting-transition temperature, T_{SC} . Hydrostatic pressure suppresses both the ferromagnetism and the superconductivity above a critical pressure of $P_c = 21$ kbar. Thus, it is not sufficient to be close to the ferromagnetic quantum critical point for superconductivity to occur in ZrZn_2 ; the compound must also be in the ferromagnetic state.

The superconductivity in ZrZn_2 has a number of remarkable features. First, it only appears to occur in high-purity single-crystal samples. Unconventional or non-s-wave forms of superconductivity generally require the superconducting coherence length $\xi = 290$ Å to be somewhat smaller than the electronic mean free path l_0 (ref. 13). Thus, in view of its sensitivity in ZrZn_2 to the quality of the sample, the superconductivity in ZrZn_2 is likely to be unconventional. Second, there is no superconducting anomaly in the specific heat. If we interpret this literally, it means that the superconducting state is strongly gapless with large portions¹⁴ of the Fermi surface, or even all of it, surviving in the superconducting state. The 'zero-field' superconducting transition in ZrZn_2 is fundamentally different to that in a conventional superconductor, because it occurs in the presence of ferromagnetism. The transition in ZrZn_2 has some similarities to the transition in a conventional superconductor for applied fields close to $\mu_0 H_{c2}$, where the superconducting anomaly is suppressed¹⁵. A second class of materials that show weak superconducting anomalies are the underdoped copper oxide¹⁶ superconductors. A key characteristic of the superconducting transition in a ferromagnet may be the absence of a strong specific heat anomaly. Third, the superconductivity in ZrZn_2 is observed only within the ferromagnetic phase, which poses the question: what is the microscopic relationship between ferromagnetism and superconductivity^{11,17,18}? We can exclude scenarios in which the superconductivity is due to inclusions of a second phase or a surface impurity, on the basis of thorough metallurgical tests and because the superconductivity and ferromagnetism disappear at the same pressure¹⁹. A macroscopic, that is, a uniform coexistence of the two states throughout the sample is consistent with the magnetic response we observe. In fact, an incomplete Meissner effect, that is, imperfect screening, is thought

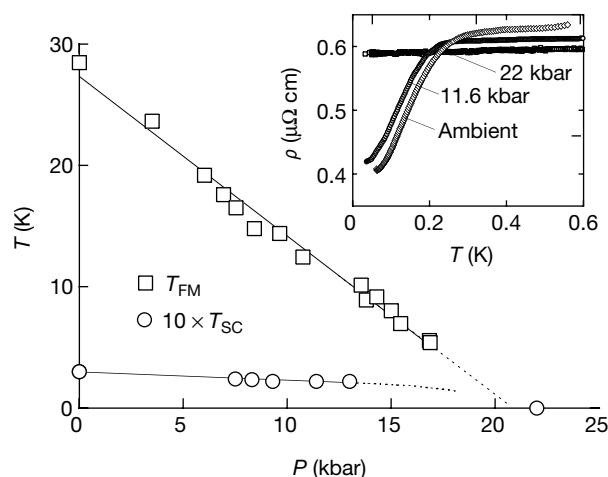


Figure 4 Pressure dependence of the ferromagnetic ordering temperature T_{FM} and superconducting ordering temperature T_{SC} . The pressure dependence of T_{FM} was determined in d.c. magnetization measurements. The pressure dependence of T_{SC} was determined from the resistivity, where typical curves are shown in the inset. Superconductivity disappears for $P > P_c = 21$ kbar, in the paramagnetic phase down to the lowest T measured, $T = 15$ mK. Note that T_{SC} for clarity is magnified by a factor of ten.

to be a signature of such a phase coexistence of superconductivity and ferromagnetism²⁰.

If metals close to a zero-temperature magnetic instability generally exhibit superconductivity, we might expect to observe this in a simple transition metal compound, as we have reported here. Our observations in ZrZn_2 contrast with materials such as ErRh_4B_4 (ref. 21) or the recently discovered $\text{RuSr}_2\text{GdCu}_2\text{O}_8$ (ref. 22) in which clearly distinguishable subsystems support either ferromagnetism or superconductivity. The presence of superconductivity throughout the entire pressure range for which ferromagnetism exists distinguishes ZrZn_2 clearly from UGe_2 (ref. 4), which is a strongly uniaxial $5f$ ferromagnet which shows a coexistence of superconductivity and ferromagnetism over a much smaller pressure range and is a member of a different class of materials. In contrast to these systems, the bands at the Fermi energy in ZrZn_2 are predominantly $\text{Zr } 4d$ (ref. 6), and the magnetism and superconductivity derive from the same $4d$ electrons. Thus far we have said little about the details of the superconductive pairing mechanism. Both the longitudinal and transverse components of the susceptibility^{5,7,8} are large for ZrZn_2 in the ferromagnetic state. This makes ZrZn_2 an excellent candidate for magnetically mediated pairing.

The most intriguing feature of the superconductivity in ZrZn_2 is that it only occurs in the presence of ferromagnetism and is hence promoted by the ferromagnetic state. This may arise naturally in scenarios where the Cooper pairs are in a parallel-spin (triplet) state, which is already favoured in the ferromagnetic state. Such behaviour could well be universal for itinerant ferromagnets in the limit of small Curie temperature and long electron mean free path. Our observations link several candidate classes of magnetically mediated superconductors in an unexpected way and suggest that fundamental notions concerning the relation of superconductivity and magnetic order need to be reexamined. □

Received 16 January; accepted 17 May 2001.

- Ginzburg, V. L. Ferromagnetic superconductors. *Sov. Phys. JETP* **4**, 153–161 (1957).
- Berk, N. F. & Schrieffer, J. R. Effect of ferromagnetic spin correlations on superconductivity. *Phys. Rev. Lett.* **17**, 433–436 (1966).
- Mathur, N. D. *et al.* Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).
- Saxena, S. S. *et al.* Superconductivity on the border of itinerant–electron ferromagnetism in UGe_2 . *Nature* **406**, 587–592 (2000).
- Matthias, B. T. & Bozorth, R. M. Ferromagnetism of a zirconium–zinc compound. *Phys. Rev.* **109**, 604 (1958).
- Jarlborg, T., Freeman, A. J. & Koelling, D. D. Self-consistent spin-polarized energy band structure and magnetism in ZrZn_2 and TiBe_2 . *J. Magn. Magn. Mater.* **23**, 291–292 (1981).
- Bernhoeft, N. R., Law, S. A., Lonzarich, G. G. & Paul, D. McK. Magnetic excitations in ZrZn_2 at low energies and long wavelengths. *Phys. Scr.* **38**, 191–193 (1988).
- van Deursen, A. P. J. *et al.* A Fermi surface study of ZrZn_2 . *J. Magn. Magn. Mater.* **54**, 1113–1114 (1986).
- Enz, C. P. & Matthias, B. T. p-state pairing and the ferromagnetism of ZrZn_2 . *Science* **201**, 828–829 (1978).
- Fay, D. & Appel, J. Coexistence of p-state superconductivity and itinerant ferromagnetism. *Phys. Rev. B* **22**, 3173–3182 (1980).
- Blagoev, K. B., Engelbrecht, J. R. & Bedell, K. S. Effect of ferromagnetic spin correlations on superconductivity in ferromagnetic metals. *Phys. Rev. Lett.* **82**, 133–136 (1998).
- Smith, T. F., Mydosh, J. A. & Wohlfarth, E. P. Destruction of ferromagnetism in ZrZn_2 at high pressure. *Phys. Rev. Lett.* **27**, 1732–1735 (1971).
- Mackenzie, A. P. *et al.* Extremely strong dependence of superconductivity on disorder in Sr_2RuO_4 . *Phys. Rev. Lett.* **80**, 161–164 (1998).
- Karchev, N. I., Blagoev, K. B., Bedell, K. S. & Littlewood, P. B. Coexistence of superconductivity and ferromagnetism in ferromagnetic metals. *Phys. Rev. Lett.* **86**, 846–849 (2001).
- Sanchez, D., Junod, A., Muller, J., Berger, H. & Levy, F. Specific heat of 2H-NbSe_2 in high magnetic fields. *Physica B* **204**, 167–175 (1995).
- Loram, J. W., Mirza, K. A., Cooper, J. R., Liang, W. Y. & Wade, J. M. Electronic specific heat of $\text{Yb}_2\text{Cu}_3\text{O}_{6.14}$ from 1.8 to 300 K. *J. Supercond.* **7**, 243–249 (1994).
- Fulde, P. & Ferrell, R. A. Superconductivity in a strong exchange field. *Phys. Rev. A* **135**, 550–563 (1964).
- Larkin, A. I. & Ovchinnikov, Y. N. Inhomogeneous state of superconductors. *Sov. Phys. JETP* **20**, 762–769 (1975).
- Huang, S. Z., Wu, M. K., Meng, R. L. & Chu, C. W. Hydrogen effect on the itinerant ferromagnets TiBe_2X and $\text{ZrZn}_{1.9}$. *Solid State Commun.* **38**, 1151–1153 (1981).
- Sonin, E. B. & Felner, I. Spontaneous vortex phase in a superconducting weak ferromagnet. *Phys. Rev. B* **57**, R14000–R14003 (1998).
- Fischer, O. & Maple, M. B. (eds) *Superconductivity in Ternary Compounds* (Springer, Berlin, 1982).

- Felner, I., Asaf, U., Levi, Y. & Millo, O. Coexistence of magnetism and superconductivity in $\text{R}_{1.4}\text{Ce}_{0.6}\text{Ru}_2\text{Sr}_2\text{O}_{10-8}$ ($\text{R}=\text{Eu}$ and Gd). *Phys. Rev. B* **55**, R3374–R3377 (1997).
- Schreurs, L. W. M. *et al.* Growth and electrical properties of ZrZn_2 single crystals. *Mater. Res. Bull.* **24**, 1141–1145 (1989).

Acknowledgements

We thank A. Griffiths for resistivity measurements, P. Pfundstein and V. Ziebat for the microprobe analysis, H. Stalzer for SQUID magnetization measurements below 1 K, and A. D. Huxley, M. B. Maple and G. Müller-Vogt for discussions. Financial support by the Deutsche Forschungsgemeinschaft (DFG), the European Science Foundation (ESF) under the FERLIN programme, and the UK Engineering and Physical Sciences Research Council (UK-EPSRC) is gratefully acknowledged.

A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning

Rafael Navarro-González*, Christopher P. McKay† & Delphine Nna Mvondo*

* Laboratorio de Química de Plasmas y Estudios Planetarios, Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Circuito Exterior, Ciudad Universitaria, Apartado Postal 70-543, México Distrito Federal 04510, México

† Space Science Division, NASA-Ames Research Center, Moffett Field, California 94035-1000, USA

Nitrogen is an essential element for life and is often the limiting nutrient for terrestrial ecosystems^{1,2}. As most nitrogen is locked in the kinetically stable form³, N_2 , in the Earth's atmosphere, processes that can fix N_2 into biologically available forms—such as nitrate and ammonia—control the supply of nitrogen for organisms. On the early Earth, nitrogen is thought to have been fixed abiotically, as nitric oxide formed during lightning discharge^{4–6}. The advent of biological nitrogen fixation suggests that at some point the demand for fixed nitrogen exceeded the supply from abiotic sources, but the timing and causes of the onset of biological nitrogen fixation remain unclear^{7–11}. Here we report an experimental simulation of nitrogen fixation by lightning over a range of Hadean (4.5–3.8 Gyr ago) and Archaean (3.8–2.5 Gyr ago) atmospheric compositions, from predominantly carbon dioxide to predominantly dinitrogen (but always without oxygen). We infer that, as atmospheric CO_2 decreased over the Archaean period, the production of nitric oxide from lightning discharge decreased by two orders of magnitude until about 2.2 Gyr. After this time, the rise in oxygen (or methane) concentrations probably initiated other abiotic sources of nitrogen. Although the temporary reduction in nitric oxide production may have lasted for only 100 Myr or less, this was potentially long enough to cause an ecological crisis that triggered the development of biological nitrogen fixation.

Because biological nitrogen fixation is energetically expensive and does not occur if adequate supplies of fixed nitrogen are available, it has been generally thought that the development of metabolic pathways to fix nitrogen arose only in response to a crisis in the supply of fixed nitrogen on the early Earth. This sudden reduction may have occurred soon after the origin of life^{7–10} as the prebiotic source of organic material was depleted by the emerging life forms. In this scenario, the abiotic sources of fixed nitrogen were unable to sustain even simple microbial ecosystems. But many modern microbially dominated ecosystems are satisfied with abiotic inflows of fixed nitrogen and do not express nitrogen fixation; thus nitrogen fixation may have arisen much later¹¹, once biological demand had

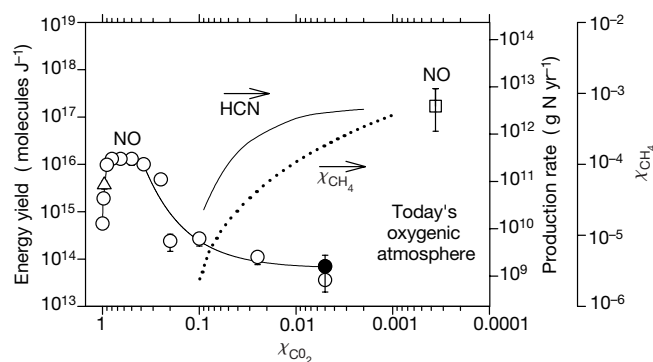


Figure 1 Variation of the nitric oxide yield as a function of the CO_2 mixing ratio in simulated lightning in primitive atmospheres. The atmospheres are composed of $\text{CO}_2\text{--N}_2$ at 1 bar, and data are shown for anhydrous (white circles) and hydrated (black circles) conditions. All samples were irradiated at 20 °C from 5 to 30 min. Immediately after irradiation, an aliquot (2 ml) was introduced with an automatic gas sampling loop into a gas chromatograph interfaced in parallel with a Fourier-transform infrared (FTIR) detector and a quadrupole mass spectrometer. Two capillary chromatographic columns were used for the separation of lightning products: PoraPlot Q (25 m $\times$ 0.32 mm internal diameter, i.d.) and Alumina/KCl (50 m $\times$ 0.32 mm i.d.). For comparison the results obtained for the simulation of the present venusian (triangle) and terrestrial (square) lightning are presented. Also shown is the possible production of HCN based on the Zahnle mechanism¹⁹ for methane mixing ratios (dotted line) needed to offset the reduced CO_2 and keep surface temperatures constant at 288.15 K (present level) based on ref. 21. Time proceeds nonlinearly from left to right.

increased owing to the development of higher plants and had exceeded the abiotic supply.

In both of these scenarios the key question is the rate of abiotic nitrogen fixation by lightning. Estimates⁴ of the lightning fixation rate on the early Earth for atmospheres containing N_2 and H_2O suggest that lightning could have provided a source of $\sim 10^{12} \text{ g N yr}^{-1}$. Calculations^{5,6} using more realistic atmospheres composed primarily of CO_2 yielded a production rate from thermodynamic equilibrium models of more than $10^{12} \text{ g N yr}^{-1}$; recycling by atmospheric photochemistry would return 25% of this to N_2 , reducing the net source of fixed nitrogen⁵. The only previous experimental determination¹² of nitrogen fixation in CO_2 -dominated atmospheres measured NO_x production in a simulated venusian atmosphere of 95.9% CO_2 , 3.97% N_2 and traces of SO_2 , Ar and CO.

The thermodynamic equilibrium models cited above^{4,5} and used in all other previous studies of nitrogen fixation by lightning on the early Earth^{11,13} are only approximate because they compute the fixation rate by assuming that the chemical reactions proceed rapidly at high temperatures and then are quenched or 'freeze out' at some specified temperature between 2,000 and 3,000 K. The equilibrium concentration at this freeze-out temperature determines the final production. To determine the nitrogen fixation accurately caused by lightning in CO_2 atmospheres, we have experimentally determined the production of NO by simulated lightning, as described in the Methods.

Figure 1 shows the variation of the NO production rate with the CO_2 mixing ratio, χ_{CO_2} . For comparison, the results obtained for the simulation of the present venusian¹² and terrestrial¹⁴ lightning are presented. Our results indicate that the energy yield of production of NO increases from $\sim 1.9 \times 10^{15} \text{ molecules J}^{-1}$ at $\chi_{\text{CO}_2} \approx 0.98$ to $\sim 1.3 \times 10^{16} \text{ molecules J}^{-1}$ at $\chi_{\text{CO}_2} \approx 0.80$. This yield is not dependent on mixing ratio down to about $\chi_{\text{CO}_2} \approx 0.50$. At lower mixing ratios the NO yield drastically decreases to $\sim 2.4 \times 10^{14} \text{ molecules J}^{-1}$ at $\chi_{\text{CO}_2} \approx 0.20$ and then slowly diminishes to $\sim 1.1 \times 10^{14} \text{ molecules J}^{-1}$ at $\chi_{\text{CO}_2} \approx 0.025$. Any further decrease

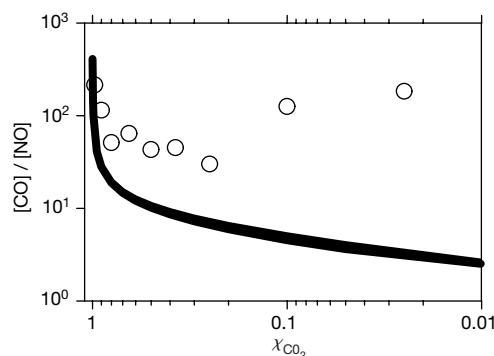


Figure 2 The ratio of NO to CO produced in laser discharge as a function of CO_2 concentration. Measurements are shown as circles. Also shown are the predictions (line) of the 'freeze-out' models for freeze-out temperatures between 2,000 and 3,000 K, the extreme range of freeze-out models.

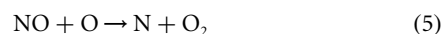
in the atmospheric CO_2 content has little impact at this point because the water vapour present in the atmosphere contributes oxygen atoms for the production of NO, stabilizing it at about $7.0 \times 10^{13} \text{ molecules J}^{-1}$. For experiments conducted at $\chi_{\text{CO}_2} \approx 0.80$, it was found that the NO production yield is not pressure-dependent. The efficient production of NO in CO_2 -dominated mixtures can be explained by the thermal dissociation of carbon dioxide leading to the formation of atomic oxygen according to reaction (1):



Most of the O produced in reaction (1) combines to form O_2 . However at temperatures above 1,500 °C, a small fraction of the O results in the formation of NO via the following reaction chain:



In competition are these reactions which destroy NO:



At low CO_2 mixing ratios, water vapour from the atmosphere contributes to the formation of atomic oxygen via the following set of reactions:



The drastic reduction in the NO production in N_2 -dominated atmospheres ($\chi_{\text{CO}_2} \leq 0.20$) is attributed to the efficient third-body conversion of atomic nitrogen into N_2 (refs 15 and 16) (see reaction (8)), transforming reaction (3) into a negligible channel.



In modern-day lightning where oxygen would be a dominant species, reactions (1) and (5) would be replaced by the following set of equations, respectively:



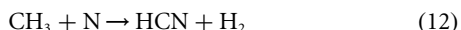
The formation of atomic oxygen in reaction (10) enhances the production of NO via reactions (2) and (10). The ratio of NO to CO in the product gases varies as expected on the basis of this reaction scheme. In our experiments, the CO to NO ratio is over 200 for CO_2

mixing ratios near unity and then drops rapidly to values between 30 and 60 for CO₂ mixing ratios between 0.8 and 0.2 (Fig. 2). Below CO₂ mixing ratios of 0.2 the CO/NO ratio climbs again to high values, reaching 180 for $\chi_{\text{CO}_2} = 0.025$. The shape of this curve for $\chi_{\text{CO}_2} > 0.2$ is as expected if the freeze-out temperature is constant. However, the increase in the CO/NO ratio at low χ_{CO_2} is not predicted by a constant freeze-out model. Freeze-out temperature variations of about 1,000 K, and different freeze-out temperatures for NO and CO would be required to explain this difference. The increasing importance of reaction (8) at low χ_{CO_2} values in our proposed chemical scheme can satisfactorily account for this increase in the CO/NO ratio.

It is unclear how the lightning flashing rate would have varied on the early Earth; however, assuming that it remained constant at $1 \times 10^{18} \text{ J yr}^{-1}$ over time¹⁷, our results (Fig. 1) imply that the annual production rate of NO was of the order of $\sim 3.0 \times 10^{11} \text{ g N}$ whenever CO₂ was the dominant atmospheric gas. This is likely to have been the case during the Hadean era when chemical evolution and the origin of life probably occurred. However, palaeosol measurements¹⁸ indicate that at 2.2 Gyr ago CO₂ levels had dropped to less than 0.04 bar. As seen from Fig. 1, this drop in CO₂ caused lightning production of NO rapidly to decrease to $\sim 2.6 \times 10^9 \text{ g N yr}^{-1}$ until a large increase in abiotic nitrogen fixation resulting either from the rise of atmospheric oxygen or production of HCN in a CH₄-rich atmosphere¹⁹. We suggest that a crisis in the supply of fixed nitrogen occurred as a result of this decrease in NO production by a factor of 100, prompting the development of biological nitrogen fixation.

Abiotic production might have increased before the rise of O₂, owing to HCN formation from CH₄ by methanogenic bacteria. Models of the early greenhouse effect suggest that biogenic methane levels increased as the CO₂ levels declined and became the dominant greenhouse gas^{20,21} with a mixing ratio of about 10^{-3} . Photochemical models¹⁹ suggest that HCN formation occurs at a significant rate when the CH₄ mixing ratio exceeds 10^{-5} for an atmosphere with 100 times the present atmospheric levels of CO₂ (see Fig. 1). Climate models²⁰ suggest that this occurred only when the CO₂ mixing ratios fell to below 0.1. Thus the nitrogen crisis would have occurred when the CO₂ mixing ratio was $0.1 \leq \chi_{\text{CO}_2} \leq 0.2$. The atmospheric CO₂ removal rate for the Precambrian is unknown; however, assuming that the CO₂ levels dropped linearly from the Hadean to early Proterozoic eons (between 2.5 and 0.6 Gyr ago), when the oldest CO₂ palaeosol records exist¹⁸, this crisis could have lasted around 10^8 yr even though it was confined to a narrow window in the atmospheric CO₂ concentration. Even if this figure were over-estimated by several orders of magnitude, a shortage of reactive nitrogen to the Archaean biosphere of more than 10^3 yr would have had dramatic ecological consequences.

The key reactions that form HCN in a CH₄-rich atmosphere are:



but their rate constants and products have not yet been studied experimentally. In fact, these reactions were considered to be similar to another reaction²²:



The rate of this reaction has been measured (it is fast) but the products have not yet been identified. Current kinetic databases¹⁶ suggest that reaction (11) does not produce HCN but rather CH + NH based on bond-energy–bond-order theory. If an NH radical is produced instead, this could lead to the production of ammonia. Therefore, the question of abiotic nitrogen fixation in CH₄-rich atmospheres is unresolved and further work is needed. If no net production of fixed nitrogen took place in CH₄-rich

atmospheres, then abiotic nitrogen fixation must have increased only with the rise of atmospheric O₂ at about 2.2 Gyr ago.

There are other indications that nitrogen fixation arose early in the history of life: (1) There is evidence of the phylogenetic antiquity of the genes for biological nitrogen fixation²³. (2) ¹⁵N/¹⁴N isotopic ratios²⁴ consistent with fixation before 3.5 Gyr ago. (3) The oxygen sensitivity of the nitrogen fixation process¹ implies an origin predating the rise of atmospheric oxygen²⁵ about 2 Gyr ago. (4) There are heterocysts formed by nitrogen-fixing cyanobacteria in fossils²⁶ aged 1.5–1.3 Gyr. (5) The antiquity of cyanobacteria is indicated by hopanoids found in 2.7-Gyr-old sediments²⁷, by morphological evidence for cyanobacteria at 3.5 Gyr ago showing microfossils that are similar to modern cyanobacteria as well as stromatolites²⁸, and by possible evidence for cyanobacteria at 3.8 Gyr ago in carbon isotopes²⁹ that show a characteristic enrichment in ¹²C.

Our results suggest that the development of biological nitrogen fixation arose in response to changes in atmospheric composition that resulted in a reduction in the production of abiotically fixed nitrogen. The biochemistry of the nitrogen-fixing process, in particular its sensitivity to oxygen, may reflect the timing of the nitrogen crisis and illustrates the co-evolution of the metabolic pathways in life and the environment of the early Earth. □

Methods

Lightning in the laboratory was simulated by a plasma generated with a pulsed Nd-YAG laser. The temperature of this plasma³⁰ has been determined to be about 20,000 K at 1 μs after initiation. The laser delivers a beam of 1.06 μm photons with an energy of up to 600 mJ per pulse in 5–7 ns at 10 Hz. A beam with 300 mJ per pulse was focused inside a closed 1-litre Pyrex flask with a plano-convex optical glass lens which is expected to have a focal aberration of about 10 μm. The power deposited into the system was determined calorimetrically. Our power meter was calibrated with the heating rate produced by passing a known electric current through a resistor. The power deposited was found to depend on the chemical composition of the gas mixture used, varying from 1.5 W for 100% CO₂ to 1.9 W for 50% CO₂ in N₂ to 2.3 W for 100% N₂. The energy dissipated per unit length in this simulated lightning was estimated to be around 10^4 J m^{-1} , in close agreement with that from natural lightning. Most experiments were conducted at 1 bar total pressure. In another series of experiments, the pressure was varied (0.2 to 1.0 bar). In some additional experiments water vapour was introduced into the reactors to achieve a 100% relative humidity. The main nitrogen-containing lightning product was determined to be nitric oxide by infrared and mass spectroscopy. The NO production rate was derived from the slope of a linear fit to the number of molecules versus energy deposited.

Received 10 November 2000; accepted 21 May 2001.

- Postgate, J. *Nitrogen Fixation* 2nd edn, 73 (Edward Arnold, London, 1987).
- Rosswall, T. in *Some Perspectives of the Major Biogeochemical Cycles* (ed. Likens, G. E.) Ch. 2 (Wiley, New York, 1981).
- Howard, J. B. & Rees, D. C. Structural basis of biological nitrogen fixation. *Chem. Rev.* **96**, 2965–2982 (1996).
- Yung, Y. L. & McElroy, M. B. Fixation of nitrogen in the prebiotic atmosphere. *Science* **203**, 1002–1004 (1979).
- Kasting, J. F. & Walker, J. C. G. Limits on oxygen concentrations in the prebiological atmosphere and the rate of abiotic fixation of nitrogen. *J. Geophys. Res.* **86**, 1147–1158 (1981).
- Kasting, J. F. Bolidic impacts and the oxidation state of carbon in the Earth's early atmosphere. *Origins Life Evol. Biosph.* **20**, 199–231 (1990).
- Cloud, P. E. Atmospheric and hydrospheric evolution on the primitive Earth. *Science* **160**, 729–736 (1968).
- Walker, J. C. G. *Evolution of the Atmosphere* (MacMillan, New York, 1977).
- Schopf, J. W. (ed.) *Earth's Earliest Biosphere: its Origin and Evolution* (Princeton Univ. Press, Princeton, 1983).
- Raven, J. A. & Yin, Z. H. The past, present and future of nitrogenous compounds in the atmosphere, and their interactions with plants. *New Phytol.* **139**, 205–219 (1998).
- Mancinelli, R. L. & McKay, C. P. The evolution of nitrogen cycling. *Origins Life Evol. Biosph.* **18**, 311–325 (1988).
- Levine, J. S. *et al.* Production of nitric oxide by lightning on Venus. *Geophys. Res. Lett.* **9**, 893–896 (1982).
- Chameides, W. L. & Walker, J. C. G. Rates of fixation by lightning of carbon and nitrogen in possible primitive atmospheres. *Origins Life* **11**, 291–302 (1981).
- Wang, Y., DeSilva, A. W., Goldenbaum, G. C. & Dickerson, R. R. Nitric oxide production by simulated lightning: Dependence on current, energy, and pressure. *J. Geophys. Res.* **103**, 19149–19159 (1998).
- Knipovich, O. M., Rubtsova, E. A. & Nekrasov, L. I. Volume recombination of nitrogen atoms in the afterglow of a condensed discharge. *Russ. J. Phys. Chem.* **62**, 867–870 (1988).
- NIST Standard Reference Database 17-2Q98 (National Institute of Standards and Technology, Gaithersburg, Maryland, 1998); also at (<http://www.nist.gov/srd/nist17.htm>).

17. Chyba, C. & Sagan, C. Electrical energy sources for organic synthesis on the early Earth. *Origins Life Evol. Biosph.* **21**, 3–17 (1991).
18. Rye, R., Kuo, P. H. & Holland, H. D. Atmospheric carbon dioxide concentrations before 2.2 billion years ago. *Nature* **378**, 603–605 (1995).
19. Zahnle, K. J. Photochemistry of methane and the formation of hydrocyanic acid (HCN) in the Earth's early atmosphere. *J. Geophys. Res.* **91**, 2819–2834 (1986).
20. Pavlov, A. A., Kasting, J. K., Brown, L. L., Rages, K. A. & Freedman, R. Greenhouse warming by CH₄ in the atmosphere of early Earth. *J. Geophys. Res.* **105**, 11981–11990 (2000).
21. Kasting, J. F., Pavlov, A. A. & Siefert, J. L. A coupled ecosystem-climate model for predicting the methane concentration in the Archean atmosphere. *Origins Life Evol. Biosph.* **31**, 271–285 (2001).
22. Yung, Y. L., Allen, M. & Pinto, J. P. Photochemistry of the atmosphere of Titan. *Astrophys. J. Suppl. Ser.* **203**, 465–506 (1984).
23. Young, J. P. W. in *Biological Nitrogen Fixation* (eds Stacey, G., Burris, R. H. & Evans, H. J.) 43–86 (Chapman & Hall, New York, 1992).
24. Beaumont, V. & Robert, F. Nitrogen isotope ratios of kerogens in Precambrian cherts: a record of the evolution of atmospheric chemistry? *Precambrian Res.* **96**, 63–82 (1999).
25. Rye, R. & Holland, H. D. Paleosols and the evolution of atmospheric oxygen: a critical review. *Am. J. Sci.* **298**, 621–672 (1998).
26. Golubic, S., Sergeev, V. N. & Knoll, A. H. Mesoproterozoic *Archaeoellipsoides*: akinetes of heterocystous cyanobacteria. *Lethaia* **28**, 285–298 (1995).
27. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).
28. Schopf, J. W. & Parker, B. M. Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* **237**, 70–73 (1987).
29. Schopf, J. W. in *Prokaryotic Development* (eds Brun, Y. V. & Shimkets, L. J.) 105–129 (American Society of Microbiology, Washington DC, 2000).
30. Jebens, D. S., Lakkaraju, H. S., McKay, C. P. & Borucki, W. J. Time resolved simulation of lightning by LIP. *Geophys. Res. Lett.* **19**, 273–276 (1992).

Acknowledgements

We acknowledge the assistance of L. Calva-Alejo. We are indebted to W. Schopf and A. Knoll for their discussions on the palaeontological evidence of nitrogen fixation. We also thank J. Kasting and A. Bar-Nun for discussions. This work was supported by grants from the National Autonomous University of Mexico, the National Council of Science and Technology of Mexico and the NASA Astrobiology programme.

Correspondence and requests for materials should be addressed to R.N.-G. (e-mail: Navarro@nuclecu.unam.mx).

Annual monsoon rains recorded by Jurassic dunes

David B. Loope*, Clinton M. Rowe* & R. Matthew Joeckel†

* Department of Geosciences, University of Nebraska, Lincoln, Nebraska 68588-0340, USA

† Conservation & Survey Division, University of Nebraska, Lincoln, Nebraska 68588-0517, USA

Pangaea, the largest landmass in the Earth's history, was nearly bisected by the Equator during the late Palaeozoic and early Mesozoic eras. Modelling experiments and stratigraphic studies have suggested that the supercontinent generated a monsoonal atmospheric circulation that led to extreme seasonality^{1–3}, but direct evidence for annual rainfall periodicity has been lacking⁴. In the Mesozoic era, about 190 million years ago, thick deposits of wind-blown sand accumulated in dunes of a vast, low-latitude desert at Pangaea's western margin^{5–7}. These deposits are now situated in the southwestern USA. Here we analyse slump masses in the annual depositional cycles within these deposits, which have been described for some outcrops of the Navajo Sandstone⁸. Twenty-four slumps, which were generated by heavy rainfall, appear within one interval representing 36 years of dune migration. We interpret the positions of 20 of these masses to indicate slumping during summer monsoon rains, with the other four having been the result of winter storms. The slumped lee faces of these Jurassic dunes therefore represent a prehistoric record of yearly rain events.

In the southwestern United States, wind-deposited sands accumulated to a thickness of 2,500 m during the 160 million years when Pangaea straddled the Equator^{6–8} (Fig. 1). These strata are the thickest and most widespread aeolian dune deposits known from the global sedimentary record. Aeolian sedimentation near the western margin of Pangaea began soon after its assembly in the Late Carboniferous period, and continued until the breakup of the supercontinent in the Late Jurassic period. The Lower Jurassic Navajo Sandstone is the thickest unit in this sedimentary succession, reaching a thickness of 700 m in southwestern Utah (17° N palaeo-latitude). Most of the Navajo Sandstone consists of thick cross-strata produced by the migration and climb of large desert dunes. The dominant winds in the area during the Triassic and Early Jurassic periods were northwesterly, whereas northeasterly flow dominated during deposition of the Middle Jurassic Entrada Sandstone⁷.

Other workers have recognized depositional cycles within the Navajo Sandstone that record an annual shift in wind direction⁸. These cycles consist of: (1) packages of 20–50 grainflow beds produced by the avalanching of dry sand as dunes migrated under the influence of northwesterly winds; and (2) wedges of wind-ripple laminae that overlie erosional surfaces produced by along-slope winds that reoriented the dune crest and moved sand obliquely up the southeast-facing slope (Fig. 2). Within packages, individual grainflow beds are separated by thin layers ('pinstripes') of finer sand deposited during short-term wind reversals (Fig. 3). Cycles in the Navajo Sandstone have been interpreted as annual increments because the distance of dune migration during a single cycle (about 1.5 m) is much too great for a day, but appropriate for a year. Longer periodicities (for example, those resulting from sunspot cycles or other multi-annual climate cycles) are highly unlikely to generate such a regular and clearly delineated record^{4,9}. Furthermore, the number of grainflow-separating pinstripes in a package (20–50) is comparable to the number of wind reversals per year at selected modern meteorological stations. In the modern subtropics and middle latitudes, surface winds reverse several times per month. At In Salah, Algeria (27° N), winds above the sand-moving threshold reversed 48 times in 1988; at Niamey, Niger (13.5° N), 15 reversals were recorded in 1984¹⁰. These data also support an annual periodicity for Navajo Sandstone depositional cycles.

Within the lower part of the Navajo Sandstone in southern Utah and northern Arizona, synsedimentary thrust faults, folds, and breccia sheets are locally common within sets of grainflow-dominated (that is, avalanche-dominated) cross-strata (see Figs 3

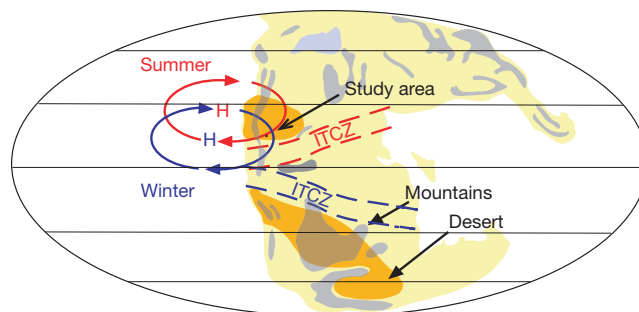


Figure 1 Pangaea during deposition of the Navajo Sandstone in the Early Jurassic. The figure shows the location of the study site (see also Supplementary Information), and the locations of deserts and mountains^{1,2}. Also shown are the suggested positions of the anticyclone near the west coast of the supercontinent (H), and of the intertropical convergence zone (ITCZ); these positions are shown for the Northern Hemisphere summer (shown red) and winter (shown blue).

and 4, and Supplementary Information). The folded and faulted strata and breccia clasts consist of grainflow and wind-ripple strata that moved down the dip of the cross-strata. The upper surfaces of these disturbed zones are sharply delineated by smooth erosional surfaces that are overlain by undeformed dune deposits (see Figs 3 and 4, and Supplementary Information). These relationships indicate that deformation pre-dated cementation, and took place on the lee face of the dunes. We interpret the faults, folds, and breccia as products of the downslope movement of cohesive sediment masses. Rain-wetted blocks of cohesive sand have been observed on many modern dune slopes, and the internal structure of the coastal dunes of Brazil and Oregon^{11–13} shows many features similar to those that we have observed in Jurassic outcrops. In the Navajo Sandstone, the masses of cohesive sand are isolated within a voluminous matrix of material emplaced by avalanches of dry, cohesionless sand. We infer that these two distinct types of mass movement record wet and dry conditions on the dune slope at different times.

The thrust slabs record movement of coherent masses of dune sand, which, in a desert, would only be likely if porewater pressure were increased. In modern dunes on the Oregon coast, mass movements of cohesive sand are generated by failure of upper dune slopes that have been oversteepened by wet grainfall—a process operative when sand-driving winds are accompanied by rain¹³. However, because the Jurassic dunes were at least 33 m high⁴, and because thrust faults are present in the wind-ripple deposits abutting lower dune slopes, it is improbable that the slide masses record oversteepening either by wet deposition at the top of slope or by undercutting at the toe. In an inland setting dominated by dry grainfall, the dune slope would have been saturated only during the very short time before water delivered by a heavy rainfall event drained into the dune. Avalanched sand typically has high porosity, sometimes exceeding 50% (ref. 14). Collapse of such loosely packed sand as the wetting front migrated downwards during a rainstorm could have rapidly increased the porewater pressure

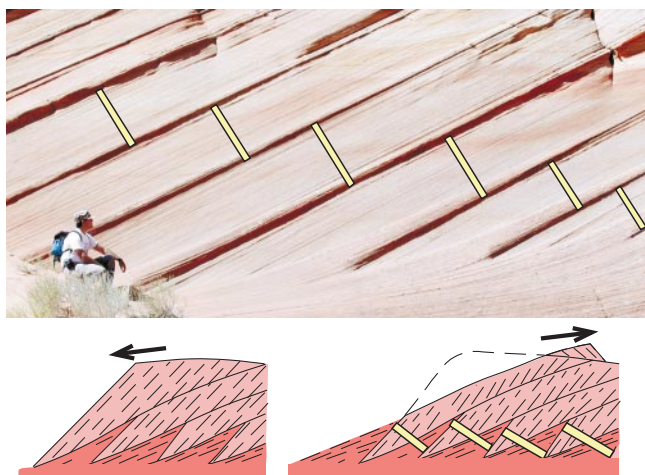


Figure 2 Annual cycles within ancient dune cross-stratification in lower Navajo Sandstone at Coyote Buttes, Arizona. Top, wedges of wind-ripple-laminated sand extend upwards from the plinth at the base of a dune to become intercalated with packages of grainflow beds deposited by dry avalanches down the slipface of the dune, which is migrating from right to left. Undeformed yearly depositional cycles (marked by yellow columns) begin at the base of one wedge and end at the erosion surface at the base of the next wedge to the left. Bottom, schematic diagram showing the effects of the seasonal shift of northwesterly winds (bottom left) to northeasterly winds (bottom right): the dune migrates under the influence of dominant northwesterly winds as dry grainflows (light red) are deposited (bottom left), whereas the seasonal wind shift (bottom right) erodes the dune top and deposits a wedge of wind-ripple-laminated sand (dark red).

along a subsurface bedding plane, triggering translational slides.

The distribution of thrust faults, folds, and breccia within annual depositional cycles enables us to determine not only the frequency of slumping events, but also their timing relative to seasonal shifts in wind direction. Although many sets of cross-strata in our field area lack disturbed sand (Fig. 2), one set (A) that records a 36-year interval of dune migration contains 24 zones of deformed strata (Fig. 5). An underlying set (B) contains 17 cycles and has 5 deformed zones. Evidence of slumping is rare in the central part of depositional cycles. If the boundary between adjacent depositional cycles is defined as the base of the wind ripple wedge (Fig. 2), then slumped

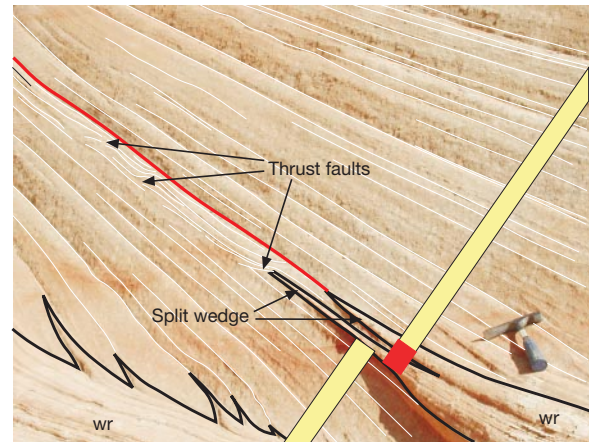


Figure 3 Parts of two successive annual depositional cycles. The cycles are shown by superimposed yellow and red bars, the positions of pinstripes (see text) are shown by white lines between grainflows, wedges of wind-rippled sand are indicated by wr, and thrust faults are arrowed. A 15-cm-thick mass of deformed sediment (red part of upper bar) lies below the erosion surface (thin red line). Rain-induced slumping followed the shift to along-slope (northeasterly) wind that deposited wind-rippled sand. Slumped sediments were buried by more wind-rippled sand before the dominant (northwesterly) wind resumed. We interpret the slumping to have taken place in summer, and the two halves of the 'split wedge' to be spring and autumn deposits. The grainflows and pinstripes, which make up more than 80% of the sediment, were deposited by the dominant winter winds.

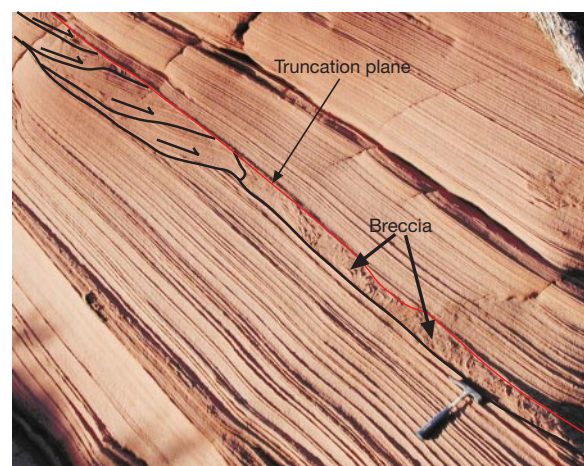


Figure 4 Slump mass and enclosing strata. The figure shows syndepositional thrust faults (arrows in upper left), and breccia that has been truncated by the erosion surface (red line) and overlain by undeformed grainflow beds. Deformation is interpreted as rain-induced slump.

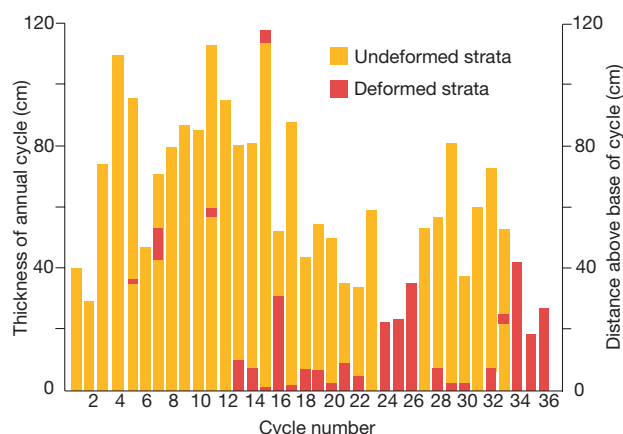


Figure 5 The properties of a continuous sequence of 36 annual cycles within 6-m-thick cross-strata of set A. Strata deformed by rain-induced slumping (red) are typically at the base or top of annual cycles, suggesting strong seasonality of major rainfall events (that is, monsoonal circulation). Thin deformed zones in mid-cycle positions (5, 7, 11 and 33) are interpreted as products of winter storms. Measured thicknesses of deformed zones were decreased by 50% to offset structural thickening. Location by Global Positioning System: 36°55' 49.4"N, 112°0' 43.5"W.

strata are clearly clustered at this surface. Of the 24 deformed zones in set A, 20 are bounded (above or below) by this surface. All 5 of the deformed zones in set B are in this position (see Supplementary Information). Both sets A and B are exposed in dip-parallel section. Only if all slumping were laterally continuous along the entire dune lee face would such a section intersect all slumps. Therefore, for this 36-year interval, we interpret slumping to have been an annual process.

We consider the dominant dune-driving winds to have been the northwesterly winter flow around the high-pressure zone at the western margin of Pangaea. The opposing winds that moved sand obliquely up the slip faces were the northeasterly trade winds of spring and autumn. Northeasterly flow at the southern edge of the dune field was followed by the arrival of heavy rains at the northern margin of the intertropical convergence zone (ITCZ). Slumps generated during these rains would have been truncated and buried by wind-ripple deposits when northeasterly flow resumed as the ITCZ migrated southwards again. The dunes would have resumed their southeastward migration in winter as the anticyclone off the west coast of Pangaea migrated southwards (Fig. 1).

The thicknesses of the deformed zones in Navajo Sandstone strata record rainfall magnitudes of individual storms. The thickest deformed zone (cycle 26, set A) measures 84 cm. Assuming a doubling of thickness during thrusting, and that the slope failure was caused by positive porewater pressure, a slab of sand about 42 cm thick would need to become saturated by water. For our calculation, we assume a porosity of 50%, residual moisture (before rainfall) of another 10%, and that failure and saturation were coincident. A minimum of 170 mm of rain would be required to saturate and destabilize such a slab.

Our observations and interpretations coincide closely with results from the only quantitative Early Jurassic climate model that is currently available³. In particular, this model shows that for the Navajo sand sea, winter winds were stronger than summer winds. It also shows that, during the summer, monsoonal flow of moist air entered the study area from the west and delivered an average of 2 mm rainfall per day. If the precipitation fell as heavy downpours from cumulonimbus clouds, such a quantity would be sufficient to generate the slumps that we have observed in the Navajo Sandstone. □

Received 15 January; accepted 2 May 2001.

- Parrish, J. T., Ziegler, A. M. & Scotese, C. R. Rainfall patterns and the distribution of coals and evaporites in the Mesozoic and Cenozoic. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **40**, 67–101 (1982).
- Kutzbach, J. E. & Gallimore, R. G. Pangaea climates: megamonsoons of the megacontinent. *J. Geophys. Res.* **94**, 3341–3357 (1989).
- Chandler, M., Rind, D. & Ruedy, R. Pangaea climate during the Early Jurassic: GCM simulations and the sedimentary record of paleoclimate. *Bull. Geol. Soc. Am.* **104**, 543–559 (1992).
- Kvale, E. P. *et al.* Evidence of seasonal precipitation in Pennsylvanian sediments of the Illinois basin. *Geology* **22**, 331–334 (1994).
- Kocurek, G. & Dott, R. H. Jr in *Mesozoic Paleogeography of West-Central United States* (eds Reynolds M. W. & Dolly, E. D.) 101–116 (Rocky Mountain Section, Society of Economic Paleontologists and Mineralogists, Denver, 1983).
- Blakey, R. C., Peterson, F. & Kocurek, G. Synthesis of late Paleozoic and Mesozoic eolian deposits of the western interior of the United States. *Sedim. Geol.* **56**, 3–125 (1988).
- Parrish, J. T. & Peterson, F. Wind directions predicted from global circulation models and wind directions determined from eolian sandstones of the western United States. *Sedim. Geol.* **56**, 261–282 (1988).
- Hunter, R. E. & Rubin, D. M. in *Eolian Sediments and Processes* (eds Brookfield, M. E. & Ahlbrandt, T. S.) 429–454 (Elsevier, Amsterdam, 1983).
- Chan, M. A. & Archer, A. W. Spectral analysis of eolian foreset periodicities: implications for Jurassic decadal-scale paleoclimatic oscillators. *Paleoclimates* **3**, 239–255 (1999).
- Loope, D. B. & Rowe, C. M. Jurassic desert storms: reactivation surfaces and wind ripple wedges that punctuate grainflow tongues within the Navajo Sandstone and record the passage of cyclonic depressions. *Geol. Soc. Am. Abstr. Progm* **32**, A308 (2000).
- Cooke, R., Warren, A. & Goudie, A. *Desert Geomorphology* Fig. 23.22, 338 (UCL Press, London, 1993).
- Bigarella, J. J., Becker, R. D. & Duarte, G. M. Coastal dune structures from Paran, Brazil. *Mar. Geol.* **7**, 5–55 (1969).
- Hunter, R. E., Richmond, B. M. & Alpha, T. R. Storm-controlled oblique dunes of the Oregon Coast. *Bull. Geol. Soc. Am.* **94**, 1450–1465 (1983).
- Hunter, R. E. Basic types of stratification in small eolian dunes. *Sedimentology* **24**, 361–387 (1977).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the Schultz Chair of Stratigraphy at the University of Nebraska.

Correspondence and requests for materials should be addressed to D.B.L. (email: dloope1@unl.edu).

Morphological and ecological complexity in early eukaryotic ecosystems

Emmanuelle J. Javaux*, Andrew H. Knoll* & Malcolm R. Walter†

* Botanical Museum, Harvard University, Cambridge, Massachusetts 02138, USA

† Australian Centre for Astrobiology, Department of Earth and Planetary Sciences, Macquarie University, New South Wales 2109, Australia

Molecular phylogeny and biogeochemistry indicate that eukaryotes differentiated early in Earth history. Sequence comparisons of small-subunit ribosomal RNA genes suggest a deep evolutionary divergence of Eukarya and Archaea; C₂₇–C₂₉ steranes (derived from sterols synthesized by eukaryotes) and strong depletion of ¹³C (a biogeochemical signature of methanogenic Archaea) in 2,700 Myr old kerogens independently place a minimum age on this split^{2,3}. Steranes, large spheroidal microfossils, and rare microfossils of possible eukaryotic origin occur in Palaeoproterozoic rocks^{4–6}. Until now, however, evidence for morphological and taxonomic diversification within the domain has generally been restricted to very late Mesoproterozoic and Neoproterozoic successions⁷. Here we show that the cytoskeletal and ecological prerequisites for eukaryotic diversification were already established in eukaryotic microorganisms fossilized nearly 1,500 Myr ago in shales of the early Mesoproterozoic Roper Group in northern Australia.

The most distinctive Roper fossils are acritarchs attributable to *Tappania plana*. *Tappania* populations consist of irregularly spheroidal organic vesicles up to 160 μm in diameter⁸ distinguished by bulbous protrusions and from zero to twenty hollow, cylindrical processes (Fig. 1a–c). The processes have closed, slightly expanded terminations and may branch dichotomously (Fig. 1b). Moreover, and in contrast to most younger acritarchs, processes are distributed irregularly and asymmetrically on the vesicle surface (Fig. 1a and c).

Acritarchs are conventionally interpreted as metabolically inert resting stages of algae, their symmetric ornamentation reflecting morphological self-organization during cyst-wall synthesis. In contrast, the irregular number and length, asymmetric distribution, and branching of processes in *Tappania* suggest an actively growing cell or germinating cyst⁹. The bulbous protrusions in some specimens further suggest vegetative reproduction through budding.

The kind of dynamic cell remodelling documented by *Tappania* morphologies does not occur in prokaryotic organisms. In eukaryotes, it is made possible by the cytoskeleton, an internal scaffolding of actin and other proteins regulated by a sophisticated network of molecular signalling pathways^{10,11}. The systematic relationships of *Tappania* are uncertain, but its distinctive morphology indicates that by 1,500 Myr ago protists already possessed the cytoskeletal

architecture and regulatory networks that characterize living protists. Thus, the cytological prerequisites for morphological radiation in eukaryotes were in place long before the taxonomic diversification documented by Neoproterozoic fossils.

Tappania was first discovered in carbonaceous shales of the Ruyang Group, northern China⁸, but beyond being demonstrably

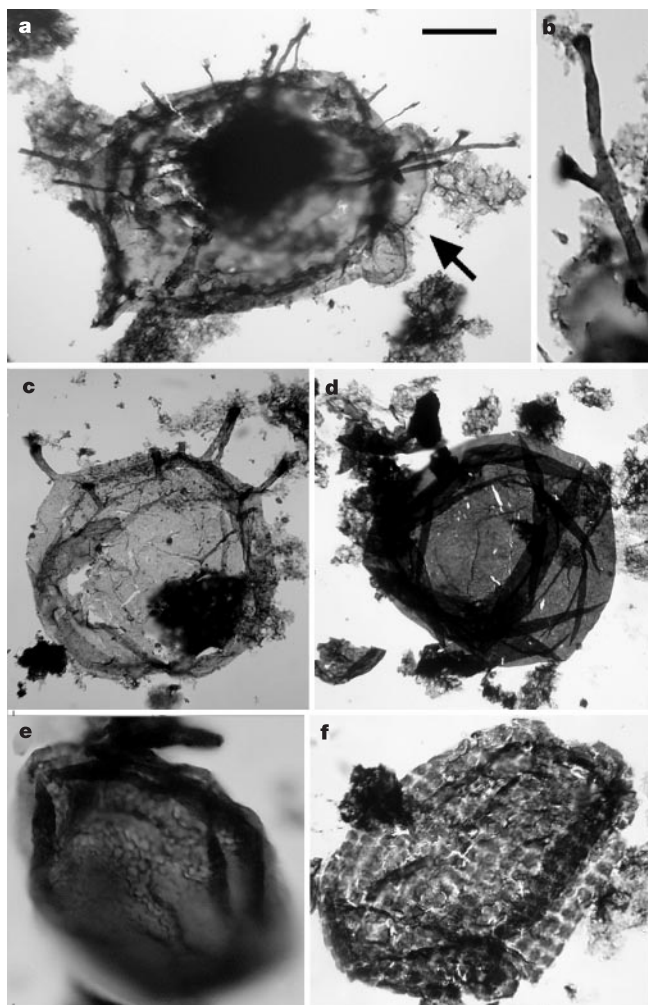


Figure 1 Protistan microfossils from the Roper Group. **a, c, e**, *Tappania plana*, showing asymmetrically distributed processes and bulbous protrusions (arrow in **a**). **b**, detail of **a**, showing dichotomously branching process. **d**, *Valeria lophostriata*. **e**, *Dictyosphaera* sp. **f**, *Satka favosa*. The scale bar in **a** is 35 μm for **a** and **c**; 10 μm for **b**; 100 μm for **d**; 15 μm for **e**; and 40 μm for **f**.

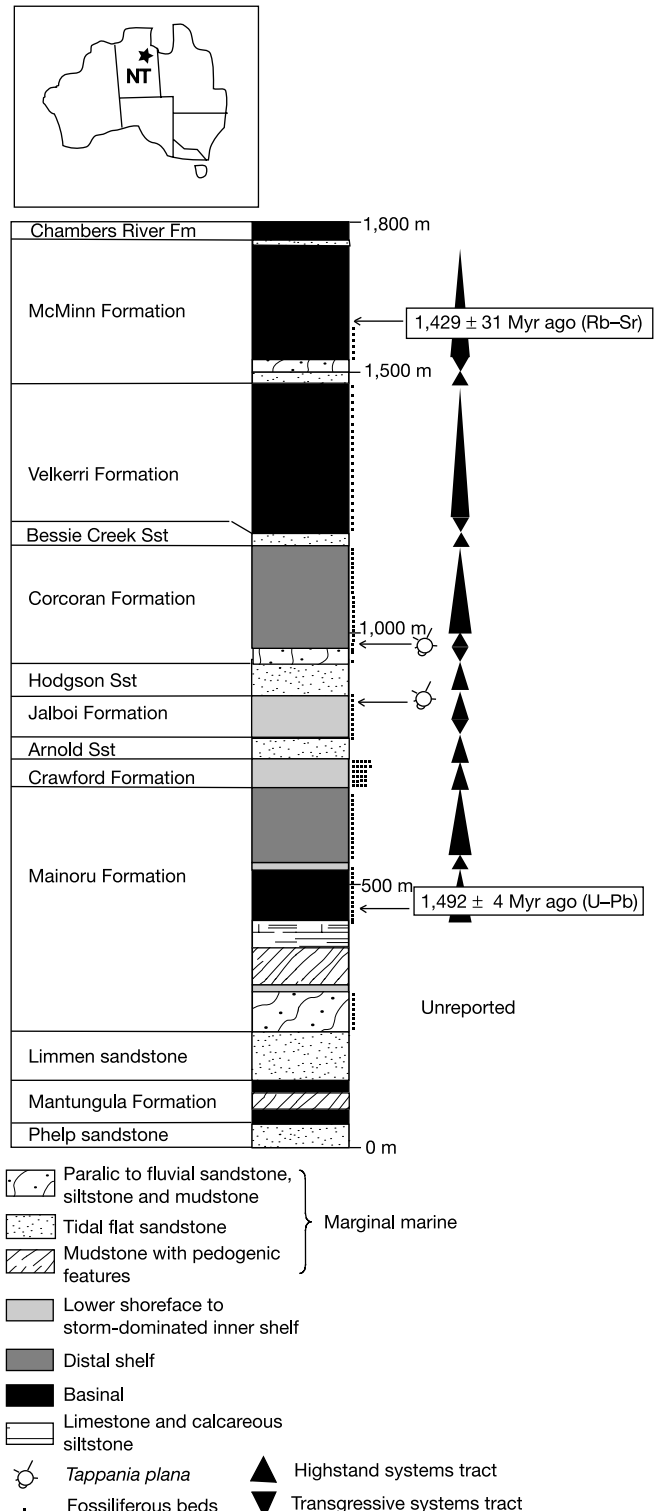


Figure 2 Location and generalized stratigraphy of the Roper Group, northern Australia. The stratigraphic distribution of facies, systems tracts and microfossils is shown, including the distinctive process-bearing protist *Tappania*. Modified from ref. 15; radiometric data from refs 13 and 14. NT, Northern Territory; Sst, Sandstone; Fm, Formation.

Mesoproterozoic, the age of the Chinese fossils is poorly constrained¹². In contrast, the new Australia populations occur in a basin that is well dated, exceptionally well characterized in terms of sedimentary architecture, and abundantly fossiliferous (Fig. 2). Thus, we are able to constrain the age, environment and ecological associations of Australian *Tappania*.

U–Pb SHRIMP analyses of zircons from an ash bed in the Mainoru Formation fix an age of $1,492 \pm 3$ Myr for early Roper deposition¹³. A $1,429 \pm 31$ Myr Rb–Sr age for illite in dolomitic siltstones near the top of the succession¹⁴ is consistent with the zircon age, albeit less reliable. Abbott and Sweet¹⁵ interpreted the stratigraphic architecture of Roper rocks in terms of six principal depositional sequences bounded by conformable surfaces in the basin interior. Sequences are highly asymmetric, consisting mostly of highstand systems tracts dominated by deep basinal to distal shelf mudstones that become more shallow upward to coastal sandstones and intercalated shales. Marine deposition is indicated by facies development and associations, which closely parallel those of younger marine successions. This is supported by distinctive sedimentary structures such as hummocky cross-stratification; evidence of reversing tidal events; and C–S ratios generally characteristic of marine basins^{15,16}.

Fossils were first discovered in Roper rocks by Peat *et al.*¹⁷, who reported leiosphaerid acritarchs, clusters of small spheroidal unicells and filamentous microfossils in shales of the McMinn Formation, near the top of the Roper succession. Our study of 126 samples from five drill cores (Urapunga 4, 5, 6; Amoco 82/3; Golden Grove 1) shows that microfossils are both abundant and unusually well preserved throughout the 2,000-m-thick group (Fig. 2).

The most abundant and species-rich Roper assemblages come from estuarine or deltaic to tide-dominated shoreline facies (Fig. 3). Coastal shales contain large (>200 µm) spheroidal microfossils, including *Leiosphaeridia* spp., the distinctively striated *Valeria lophostriata* (Fig. 1d), and thick-walled vesicles more than a millimetre in diameter. Conventional (50–100 µm) leiosphaerids are also abundant, as is *Satka favosa*, a spheroidal acritarch with walls of interlocking rectangular panels (Fig. 1f). Although species richness is highest in coastal samples, these assemblages also exhibit strong concentration of dominance in one or two species (Fig. 3); most of the diversity is contributed by rare taxa.

Shales in lower-shoreface to storm-dominated inner-shelf facies preserve a lower abundance and diversity of sphaeromorphic

acritarchs, common *Satka squamifera* (distinguished by rounded wall textures imparted by included cell packets), and large (up to 150 µm in cross-sectional diameter) striated tubes of uncertain, but probably eukaryotic, affinities. (A few sulphur-oxidizing bacteria can produce comparably large sheaths, but none is known to produce thick, regularly striated walls comparable to the Roper fossils.) More distal shelf shales contain still fewer fossils. *Tappania* populations are restricted to this facies, where they occur with filamentous, probably cyanobacterial, sheaths and scattered spheroidal acritarchs. This may explain the apparent absence of *Tappania* in well studied Mesoproterozoic successions from the Urals and Siberia, which are dominated by inshore depositional facies^{18,19}. Chinese *Tappania* populations also occur in distal shelf shales²⁰.

Most basinal mudstones contain abundant organic debris but few well preserved microfossils. A few samples, however, contain monospecific assemblages of *Stictosphaeridium* sp., 40–70 µm spheroidal acritarchs with thin, chagrinat walls.

The onshore–offshore pattern (Fig. 3) of decreasing abundance, declining diversity, and changing dominance among fossils of probable eukaryotic origin is observed consistently throughout our large sample set. Younger successions commonly show maximum eukaryotic diversity in inner shelf settings, with decreasing diversity offshore^{21–24}. Rarely, however, do abundance and diversity peak in marginal marine facies, as observed in the Roper Group. An explanation for this unusual palaeoenvironmental pattern may lie in chemistry of Mesoproterozoic oceans, where biologically important trace elements (especially Mo, required for nitrate assimilation) were possibly scarce in offshore waters distant from continental runoff^{25–26}. Regardless of the explanation for fossil distributions in the Roper Group, the simple observation of strong facies associations among taxa shows that 1,500 Myr ago natural selection had long since begun to distribute protistan populations among physical habitats—contributing to the rise of biological diversity (and providing tools for palaeoenvironmental reconstruction).

The discovery of distinctive microfossils in the well dated Roper Group and its Chinese counterpart raises the prospect of a biostratigraphic characterization of early Mesoproterozoic sedimentary rocks. Process-bearing microfossils also occur in younger Mesoproterozoic successions, but they are distinct from those observed in Roper and Ruyang shales^{27–28}. We predict that the Mesoproterozoic era will prove to be divisible into at least two broad intervals based on microfossil assemblages and carbon isotope profiles, although

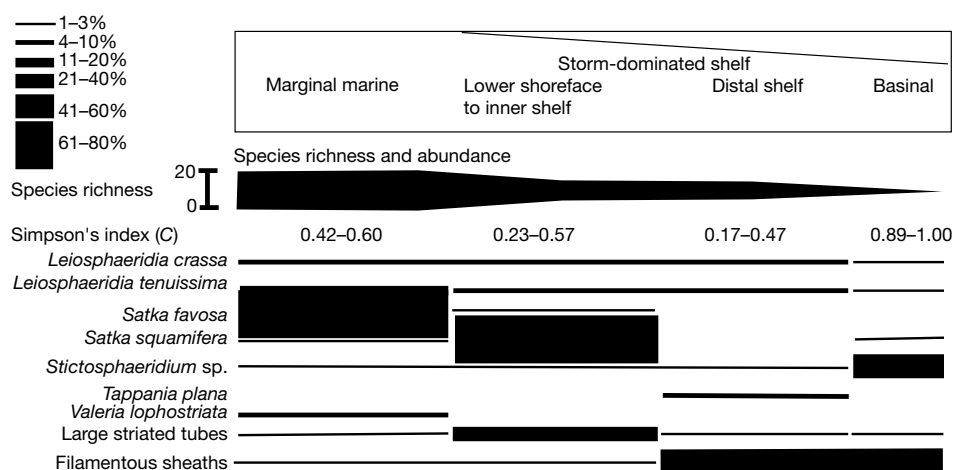


Figure 3 Palaeoenvironmental distribution of Roper fossils, showing relationships among physical environment, fossil abundance, and taxonomic diversity. Of the 126 samples used in this study, 13 come from marginal marine facies, 26 from lower-shoreface to storm-dominated inner-shelf facies, 30 from middle to distal shelf facies, and 57 from

basinal facies. Species richness and Simpson's index of dominance concentration (*C*) are based on size-standardized (*N* = 300) counts of fossils within samples. Abundance was estimated using microfossil density within samples. Legend in upper left indicates percentage contribution of listed taxa to total fossil assemblage in each environment.

the bio- and chemostratigraphic breakpoints may not coincide precisely. The restricted facies distribution of distinctive early Mesoproterozoic acritarchs, however, suggests that biostratigraphic confidence will necessarily be tied to careful palaeoenvironmental characterization.

Process-bearing microfossils were once thought to be restricted to Phanerozoic strata. Two decades ago, their range was extended to the beginning of the Neoproterozoic era. Now, it is clear that process-bearing microfossils have been present for nearly half of life's recorded history. The Roper Group provides an unusually clear window on aspects of biology in the early Mesoproterozoic oceans, demonstrating that 1,500 Myr ago marine protists included cytologically sophisticated organisms that lived in ecologically differentiated communities. □

Received 2 February; accepted 18 May 2001.

1. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579 (1990).
2. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. Archean molecular fossils and the early rise of eukaryotes. *Science* **285**, 1033–1036 (1999).
3. Hayes, J. M. in *Early Life on Earth* (ed. Bengtson, S.) 220–236 (Columbia Univ. Press, New York, 1994).
4. Zhang, Z. Clastic facies microfossils from the Chaunlinguo Formation (1800 Ma) near Jixian, North China. *J. Micropalaeontol.* **5**, 9–16 (1986).
5. Summons, R. E., Powell, T. G. & Boreham, C. J. Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, Northern Australia. III. Composition of extractable hydrocarbons. *Geochim. Cosmochim. Acta* **51**, 3075–3082 (1988).
6. Han, T.-M. & Runnegar, B. Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee Iron Formation, Michigan. *Science* **257**, 232–235 (1992).
7. Knoll, A. H. in *Palynology: Principles and Applications* (eds Jansonius, J. & McGregor, D. C.) Vol. I 51–80 (American Association of Stratigraphic Palynologists Foundation, Tulsa, Oklahoma, 1996).
8. Yin, L. Acanthomorphic acritarchs from Meso-Neoproterozoic shales of the Ruyang Group, Shanxi, China. *Rev. Palaeobot. Palynol.* **98**, 15–25 (1998).
9. Butterfield, N. J., Knoll, A. H. & Swett, K. Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Fossils Strata* **34**, 1–84 (1994).
10. Jansen, R.-P. RNA-cytoskeletal associations. *FASEB J.* **13**, 455–466 (1999).
11. Schmidt, A. & Hall, M. N. Signalling the actin cytoskeleton. *Annu. Rev. Cell Dev. Biol.* **14**, 305–338 (1998).
12. Xiao, S., Knoll, A. H., Kaufman, A. J., Yin, L. & Zhang, Y. Neoproterozoic fossils in Mesoproterozoic rocks? Chemostratigraphic resolution of a biostratigraphic conundrum from the North China Platform. *Precambrian Res.* **84**, 197–220 (1997).
13. Jackson, M. J., Sweet, I. P., Page, R. W. & Bradshaw, B. E. in *Integrated Basin Analysis of the Isa Superbasin Using Seismic, Well-log, and Geopotential Data: an Evaluation of the Economic Potential of the Northern Lawn Hill Platform* (eds Bradshaw, B. E. & Scott, D. L.) (Australian Geological Survey Organisation Record 1999/19, Canberra, Australia, 1999).
14. Kralik, M. Rb–Sr age determinations on Precambrian carbonate rocks of the Carpentarian McArthur Basin, Northern Territory, Australia. *Precambrian Res.* **18**, 157–170 (1982).
15. Abbott, S. T. & Sweet, L. P. Tectonic control on third-order sequences in a siliciclastic ramp-style basin: an example from the Roper Superbasin (Mesoproterozoic), northern Australia. *Aust. J. Earth Sci.* **47**, 637–657 (2000).
16. Jackson, M. J. & Raiswell, R. Sedimentology and carbon-sulfur geochemistry of the Velkerri Formation, a mid-Proterozoic potential oil source in northern Australia. *Precambrian Res.* **54**, 81–108 (1991).
17. Peat, C. R., Muir, M. D., Plumb, K. A., McKirdy, D. M. & Norvick, M. S. Proterozoic microfossils from the Roper Group, Northern Territory. *Bur. Min. Res. J. Austr. Geol. Geophys.* **3**, 1–17 (1978).
18. Jankauskas, T. V. *Precambrian microfossils from the USSR* 5–33 (Nauka, Leningrad, 1989).
19. Veis, A. F. & Vorobyeva, N. G. Riphean and Vendian microfossils of the Anabar Uplift. *Isvest. Akad. Nauk. USSR Ser. Geol.* **8**, 114–130 (1992).
20. Zhou, H. et al. *Sequence Stratigraphic Research in Meso- to Neoproterozoic Successions of the Southern North China Platform* 5–36 (Geological Publishing House, Beijing, 1999).
21. Jacobsen, S. R. Acritarchs as paleoenvironmental indicators in Middle and Upper Ordovician rocks from Kentucky, Ohio, and New York. *J. Paleontol.* **53**, 1197–1212 (1979).
22. Vidal, G. & Nystuen, J. P. Micropaleontology, depositional environment, and biostratigraphy of the Upper Proterozoic Hedmark Group, southern Norway. *Am. J. Sci.* **A 290**, 170–211 (1990).
23. Butterfield, N. J. & Chandler, F. W. Palaeoenvironmental distribution of Proterozoic microfossils, with an example from the Agu Bay Formation, Baffin Island. *Palaeontology* **35**, 943–957 (1992).
24. Petrov, P. Yu. & Veis, A. F. Facial-ecological structure of the Derevnya Formation microbiota: Upper Riphean, Turukhansk Uplift, Siberia. *Stratigr. Geol. Correl.* **3**, 435–4605 (1995).
25. Canfield, D. E. A new model for Proterozoic ocean chemistry. *Nature* **396**, 450–453 (1998).
26. Anbar, A. D. & Knoll, A. H. Trace metal limitation of primary production 1.85–1.25 Ga. 1999 American Geophysical Union Fall Meeting (<http://www.agu.org/meetings/fm99top.html>) (1999).
27. German, T. N. *Organic World One Billion Years Ago* 1–50 (Leningrad, Nauka, 1990).
28. Samuelsson, J., Dawes, P. R. & Vidal, G. Acid-resistant palynomorphs from the Proterozoic Thule Group, northwest Greenland. *Precambrian Res.* **96**, 1–23 (1999).

Acknowledgements

This paper is a contribution to the Australian Geological Survey Organization's NABRE project. We thank J. Jackson, P. Southgate and other members of the NABRE team for access to unpublished observations, helpful discussions, and advice on sampling. Core library staff in Canberra (AGSO) and Darwin (Northern Territory Geological Survey) greatly facilitated sample collection. Y. Leiming provided information on Chinese

stratigraphy. Research supported in part by a grant from NASA, the NASA Astrobiology Institute, the Australia Research Council, and Macquarie University.

Correspondence and requests for materials should be addressed to A.H.K. (e-mail: aknoll@oeb.harvard.edu).

The impact of ultraviolet radiation on the vertical distribution of zooplankton of the genus *Daphnia*

Stephan C. Rhode, Markus Pawlowski & Ralph Tollrian

Ludwig-Maximilians-Universität München, Department of Ecology, Zoological Institute, Karlstraße 25, D-80333 München, Germany

The vertical migration of zooplankton into lower and darker water strata by day is generally explained by the avoidance of visually orienting predators, mainly fish^{1–4}; however, it is unclear why daily zooplankton migration has been maintained in fishless areas⁵. In addition to predation, ultraviolet radiation—a hazardous factor for zooplankton in the surface layers of marine and freshwater environments^{6–8}—has been suspected as a possible cause of daytime downward migration⁹. Here we test this hypothesis by studying several *Daphnia* species, both in a controlled laboratory system and under natural sunlight in an outdoor system. We selected *Daphnia* species that differed in their pigmentation as both melanin and carotenoids have been shown to protect *Daphnia* from ultraviolet light^{10,11}. All *Daphnia* species escaped into significantly deeper water layers under ultraviolet radiation. The extent to which the daphnids responded to this radiation was inversely linked to their pigmentation, which reduced ultraviolet transmission. These results suggest that ultraviolet avoidance is an additional factor in explaining daytime downward migration. Synergistic benefits might have shaped the evolution of this complex behaviour.

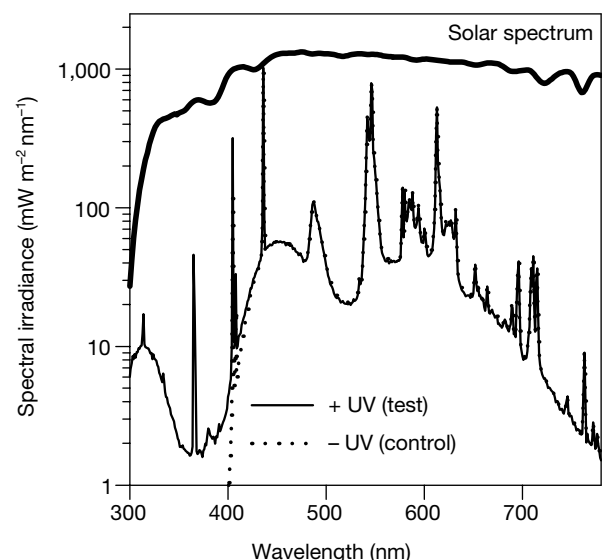


Figure 1 Spectral irradiation in mesocosms with (test) and without (control) ultraviolet radiation. Ultraviolet radiation above the water surface was lower than on a cloudless September day at noon in 1995 in Munich, when we performed the experiments under natural solar irradiation (solar spectrum).

Diel vertical migration (DVM) in zooplankton is an impressive phenomenon, involving tons of biomass changing its vertical position in a daily rhythm, both in marine and freshwater environments, with profound impact on all trophic levels. This phenomenon was noticed by the earliest scientists in this field and is reflected in the implicit knowledge of fishermen over the centuries. In cases of predator-induced DVM, zooplankton avoid 'light-dependant mortality'¹² from visually feeding predators by staying in deeper, darker water layers during the day¹³. Changes in light intensity trigger upward and downward DVM of zooplankton¹⁴; chemical or physical stimuli stemming from predators can also induce DVM behaviour¹⁵. However, migrating zooplankton normally have to cross temperature gradients in the natural water column, and the smaller growth rates at cooler temperatures in lower water strata is a major cost¹⁶. Thus, additional benefits to predator avoidance may be necessary to explain the evolution and persistence of this complex and costly behaviour. Furthermore, there are vertically migrating zooplankton communities even in the absence of fish predators in arctic lakes⁵, a result that has yet to be explained.

Another harmful factor for zooplankton is ultraviolet-B radiation^{6–8}, which may penetrate to significant depth. In clear marine water 10% of surface ultraviolet radiation penetrated down to 25 m (ref. 17), and in low dissolved organic carbon (DOC) freshwater lakes a maximum penetration depth for 1% of the surface level of UV-B was measured at 33 m (ref. 18). For *Daphnia magna*, ultraviolet receptors¹⁹ and the ability to see ultraviolet radiation²⁰ have been shown. Given that ultraviolet radiation is a selective ecological factor, a similar avoidance reaction could be expected. To test this hypothesis we performed depth-selection experiments with daphnids. Our goal was to test if the ultraviolet perception, enabling for a negative phototactic short-term response to ultraviolet radiation^{21–23}, would result in appropriate depth selection of differently pigmented daphnids when exposed to ultraviolet radiation. We performed experiments under both natural and artificial sunlight. We measured the stabilized vertical distribution of pigmented and unpigmented daphnids in experimental mesocosms. Under natural sunlight we found a significant deeper migration when the daphnids were exposed to ultraviolet irradiation (Mann–Whitney *U*-test, $P \leq 0.0001$; Table 1). We verified our results in laboratory experiments. Our experimental ultraviolet intensity of 0.505 mW m^{-2}

Table 1 Median depth distributions of adult and juvenile *Daphnia* under simulated and natural sunlight

	Median of depth distribution (m)					
	Ultraviolet		Control		Difference	
	Adults	Juveniles	Adults	Juveniles	Adults	Juveniles
Under laboratory irradiation						
<i>D. cucullata</i>	0.65	0.50	0.20	0.15	0.45	0.35
<i>D. pulex</i> (unpigmented)	0.85	0.85	0.15	0.10	0.70	0.75
<i>D. pulex</i> (melanized)	0.15	0.20	0.10	0.10	0.05	0.10
<i>D. rosea</i>	0.20	0.30	0.10	0.10	0.10	0.20
Under solar irradiation						
<i>D. pulex</i> (unpigmented)	0.85	–	0.45	–	0.40	–
<i>D. pulex</i> (melanized)	0.60	–	0.55	–	0.05	–

(285–400 nm) falls below the natural ultraviolet radiation of temperate cloudless spring, summer and autumn days at noon (Fig. 1). By applying steep temperature and irradiation gradients (Fig. 2e, f), we chose experimental conditions that allowed detection of graded responses in our small-scale set-up. All of our experiments in mesocosms showed a significant downward migration of ultraviolet-exposed daphnids into the colder part of the mesocosms (Mann–Whitney *U*-test, $P \leq 0.0001$; Fig. 2 and Table 1). The daphnids avoided ultraviolet radiation despite the anticipated cost.

Melanin pigmentation as well as carotenoids increase the ultraviolet tolerance of planktonic species^{10,11}. To minimize the cost of colder temperatures in deeper water layers, a less pronounced downward response to ultraviolet light would be predicted for zooplankton with a higher ultraviolet tolerance. This was confirmed by the fact that the response to ultraviolet light was weaker in both of the pigmented forms (Fig. 2, Table 1). Under ultraviolet radiation pigmented *Daphnia pulex* and *Daphnia rosea* stayed closer to the surface than unpigmented *D. pulex* and *Daphnia cucullata*. This is reflected in a smaller difference in median depth between the control groups and the ultraviolet treatments for either of the pigmented clones. Under natural solar radiation in outdoor experiments our results confirmed a more pronounced downward response of adult unpigmented *D. pulex* compared with melanized *D. pulex* (Table 1). Melanin in the carapace effectively reduces

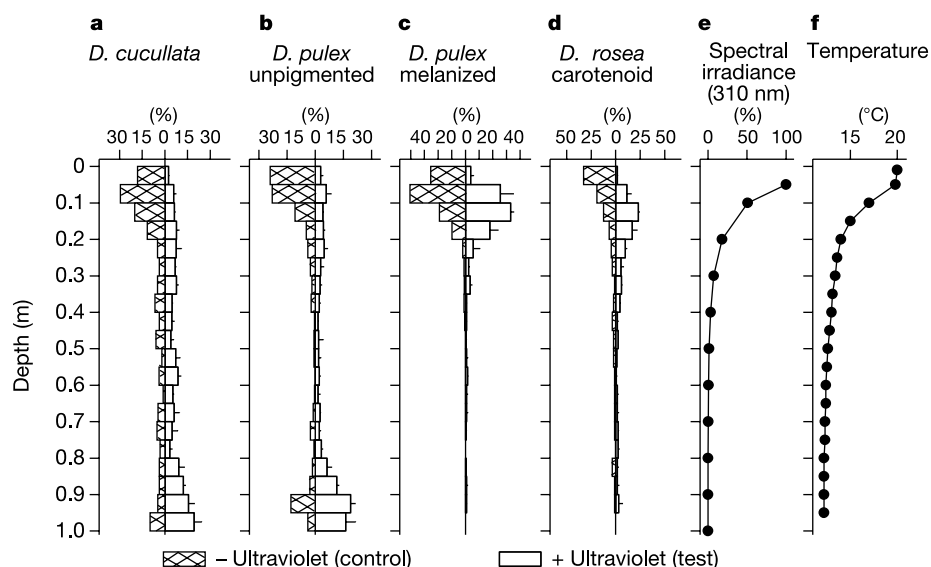


Figure 2 Results from laboratory experiments. **a–d**, Vertical distribution of unpigmented *D. cucullata* (**a**) and *D. pulex* (**b**); melanized *D. pulex* (**c**); and *D. rosea* with carotenoids (**d**), in mesocosms (height 1 m; diameter 46 mm). **e, f**, We measured the vertical gradient of radiation (**e**) and the vertical temperature gradient (**f**). Data represent means of three

replicates (± 1 s.d.) taken as the mean of five repeated measurements. Measured values of depth distribution in all experiments showed a significantly deeper position of daphnids in ultraviolet treatments (Mann–Whitney *U*-test; $P < 0.0001$).

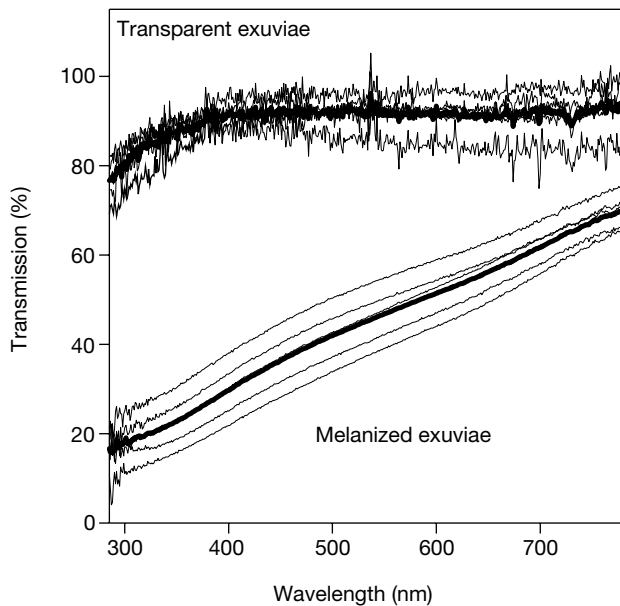


Figure 3 Spectral transmission of exuviae from transparent and melanized *D. pulex* measured in 1-nm intervals from 280 to 780 nm. Thick lines represent means. Melanin pigmentation effectively reduced transmission of ultraviolet irradiation.

ultraviolet transmission (Fig. 3). The downward migration of both pigmented and unpigmented daphnids during the day indicates that ultraviolet radiation is a proximate factor in depth selection. The stronger response of unpigmented daphnids apparently compensates for a weaker ultraviolet tolerance compared with that of their pigmented congeners. This implies that ultraviolet radiation is not only a proximate but, at the same time, is also an ultimate factor for the migration of zooplankton.

If ultraviolet radiation is both an ultimate and a proximate factor it could have had an initial role in the evolution of the predator-induced daytime sinking in DVM²⁴. Perhaps the benefit for populations whose migration was originally ultraviolet-induced would have increased with the abundance of visually orienting predators. An extended migration amplitude would also have further contributed to the benefit of vertical migration, and then been evolutionarily established as predator-induced DVM in zooplankton. The depth necessary to avoid visually orienting predators²⁵ generally exceeds the depth to which damaging ultraviolet intensities penetrate⁷, as visible light penetrates deeper into natural waters than does ultraviolet radiation²⁶, and fish are able to forage at very low light levels^{27,28}. Owing to the abundance of fish in many lakes, ultraviolet protection would already be achieved by a predator-induced DVM. Nevertheless, if *Daphnia* can avoid being circulated to the surface by turbulent mixing in the epilimnion (circulating surface water), a migration into the hypolimnion (non-circulating, lower water layer) might even be advantageous in lakes where ultraviolet penetration is reduced by higher contents of DOC¹⁸. It follows that the impact of natural and anthropogenically increased ultraviolet radiation²⁹ on zooplankton DVM should have an involvement in aquatic environments with temporally or spatially low fish abundance, such as in high mountain lakes, arctic regions and in many marine areas. On the basis of our results we propose that pigmentation and migration are two alternative strategies to avoid harmful ultraviolet radiation and we expect pigmentation, as it increases visibility and thus susceptibility to visual predation, to be found in lakes without fish, and migration to be found in lakes with fish. Our study shows that depth-selection behaviour can be a multi-causal phenomenon. The additive benefits possibly increase the adaptive value and thus facilitate the evolution and persistence of such a complex behaviour. □

Methods

Cultures

Cultures of unpigmented *D. cucullata* originated from Lake Thalersee (Bavaria, Germany) and *D. rosea* with carotenoid body pigmentation originated from a fishless pond in the Bavarian alps (Germany). Two *D. pulex* clones, one an unpigmented clone and the other with a melanized carapace, were isolated from Churchill in arctic Canada.

Mesocosm experiments

For the experiments we constructed Perspex tube mesocosms (height 1 m; diameter 46 mm). For the laboratory experiments the mesocosms were irradiated from above by one Phillips TL12/40W ultraviolet-B fluorescent tube, two Osram L36/W/11 Lumilux daylight and two Salina ES Standard F36W/129ST tubes giving a light spectrum that covered the ultraviolet and visible bandwidth of natural sunlight. Light was scattered by an opalescent ultraviolet-permeable Perspex sheet. We used acetate foil to eliminate ultraviolet-C radiation of the Phillips ultraviolet-B fluorescent tube. Ultraviolet radiation was excluded from controls by Mylar foil. Three control and three ultraviolet treatments were positioned in an alternating order and screened off from the others to prevent systematic errors. The total irradiation difference between ultraviolet and control treatments (wavelength (λ) > 400 nm) was less than 2%. Spectral irradiance was measured with a calibrated Bentham spectral radiometer. For the experiments with natural sunlight the system was transported to the roof of the institute, where it had been installed in a dark container allowing light to penetrate the Perspex tubes only from above.

The mesocosms were filled with filtered (0.45 μ m) water from the eutrophic Lake Klostersee (Germany), which had been stored for more than a week to exclude the possible effects of algae and kairomones. K_{d310} (the diffuse attenuation coefficient for wavelength 310 nm) in Lake Klostersee (measured in the lake) was 4.33 m^{-1} . Owing to the experimental geometry the attenuation was higher in the experimental columns: $K_{d310}(exp) \sim 10 m^{-1}$.

To obtain vertical gradients of radiation (Fig. 2e) we constructed Perspex tubes of each depth, sealed the bottom with ultraviolet-permeable acetate foil and measured the ultraviolet radiation that penetrated each water column. The system was cooled to 12 °C and heated from above by lamp irradiation to 20 °C, which resulted in a steep vertical temperature gradient (Fig. 2f). We optimized the system in a series of previous experiments to obtain not only directional, but also graded responses. We added 20 light-acclimated (20 $mW m^{-2}$) daphnids to each mesocosm. For the experiments with juveniles we used 3–4-day-old (3rd to 4th instar) daphnids; for the experiments with adults we used 10-day-old daphnids. On the basis of the results of pilot experiments, we counted the individuals after 2-h acclimation of the daphnids to the light (both with and without continuous ultraviolet radiation) and temperature conditions. We verified stabilization of the behaviour by counting five times (repeated measurement, time interval of 20 min). No significant shift of the median depth was detected during the 80 min of the measurement. For statistical analysis we compared the within treatment variation between the replicates with Kruskal–Wallis tests. Owing to the homogeneous conditions in our set-up we found no significant difference between replicates. This allowed us to pool the data and to compare the treatments with Mann–Whitney U-tests.

Measuring the protective effect of pigmentation

To verify the protective effect of melanin, we measured ultraviolet transmission directly at the exuviae. Five freshly moulted exuviae from transparent and melanized *D. pulex* were washed in H₂O-bidest and transferred on quartz glass coverslips. The centre of one side of the exuviae was positioned over a 100- μ m diaphragm. The spectral irradiation of a 100-W quartz halogen calibration light bulb transmitting the cover slide with and without an exuviae was measured by a Bentham spectral radiometer between 280 nm and 780 nm at 1-nm steps. Transmission was calculated by the quotient of the spectral irradiance of an empty cover slide $\omega_{ex}(\lambda)$, and the spectral irradiance of a cover slide with exuviae $\omega_{ex}(\lambda)$: $T(\lambda) = \omega_{ex}(\lambda)/\omega_{ex}(\lambda)$.

Received 8 March; accepted 24 April 2001.

- Zaret, T. M. & Suffern, J. S. Vertical migration in zooplankton as a predator avoidance mechanism. *Limnol. Oceanogr.* **21**, 804–813 (1976).
- Stich, H. B. & Lampert, W. Predator evasion as an explanation of diurnal vertical migration by zooplankton. *Nature* **293**, 396–398 (1981).
- Neill, W. E. Induced vertical migration in copepods as a defence against invertebrate predation. *Nature* **345**, 524–526 (1990).
- De Meester, L., Weider, L. J. & Tollrian, R. Alternative antipredator defences and genetic polymorphism in a pelagic predator–prey system. *Nature* **378**, 483–485 (1995).
- Williamson, C. E. *et al.* Ultraviolet radiation and zooplankton community structure following deglaciation in Glacier Bay, Alaska. *Ecology* (in the press).
- Calkins, J. & Thordarottir, T. The ecological significance of solar UV radiation on aquatic organisms. *Nature* **283**, 563–566 (1980).
- Williamson, C. E., Zagarese, H. E., Schulze, P. C., Hargreaves, B. R. & Seva, J. The impact of short-term exposure to UV-B radiation on zooplankton communities in north temperate lakes. *J. Plankton Res.* **16**, 205–218 (1994).
- Siebeck, O. & Böhm, U. Challenges for an appraisal of UV-B effects upon planktonic crustaceans under natural radiation conditions with a non-migrating (*Daphnia pulex obtusa*) and a migrating cladoceran (*Daphnia galeata*). *Arch. Hydrobiol. Beih. Ergebn. Limnol.* **43**, 197–206 (1994).
- Kerfoot, W. C. Adaptive value of vertical migration: comments on the predator hypothesis and some alternatives. *Contr. Mar. Sci. Suppl.* **27**, 91–113 (1985).
- Hairton, N. G. Jr. Photoprotection by carotenoid pigments in the copepod *Diaptomus nevadensis*. *Proc. Natl Acad. Sci. USA* **73**, 971–974 (1976).

11. Hebert, P. D. N. & Emery, C. J. The adaptive significance of cuticular pigmentation in *Daphnia*. *Funct. Ecol.* **4**, 703–710 (1990).
12. Lampert, W. The adaptive significance of diel vertical migration of zooplankton. *Funct. Ecol.* **3**, 21–27 (1989).
13. De Meester, L., Dawidowicz, P., van Gool, E. & Loose, C. J. in *The Ecology and Evolution of Inducible Defenses* (eds. Tollrian, R. & Harvell, C. D.) 160–176 (Princeton Univ. Press, Princeton, New Jersey, 1999).
14. Ringelberg, J. The positively phototactic reaction of *Daphnia magna* Strauss: a contribution to the understanding of diurnal vertical migration. *Neth. J. Sea Res.* **2**, 319–406 (1964).
15. Bollens, S. M., Frost, B. W. & Cordell, J. R. Chemical, mechanical and visual cues in the vertical migration behavior of the marine planktonic copepod *Acartia hudsonica*. *J. Plankton Res.* **16**, 555–564 (1994).
16. Loose, C. J. & Dawidowicz, P. Trade-offs in diel vertical migration by zooplankton: The costs of predator avoidance. *Ecology* **75**, 2255–2263 (1994).
17. Fleischmann, E. M. The measurement and penetration of ultraviolet radiation into tropical marine water. *Limnol. Oceanogr.* **34**, 1623–1629 (1989).
18. Morris, D. P. *et al.* The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon. *Limnol. Oceanogr.* **40**, 1381–1391 (1995).
19. Smith, K. C. & Macagno, E. R. UV photoreception in the compound eye of *Daphnia magna* (Crustacea, Branchiopoda). A fourth spectral class in single ommatidia. *J. Comp. Physiol. A* **166**, 597–606 (1990).
20. Merker, E. Sehen die Daphnien ultraviolettes Licht? *Zool. Jahrb. Abt. Allg. Zool. Physiol. Tiere* **48**, 277–348 (1930).
21. Hessen, D. O. *Daphnia* responses to UV-light. *Arch. Hydrobiol. Beih.* **43**, 85–195 (1994).
22. Storz, U. C. & Paul, R. J. Phototaxis in water fleas (*Daphnia magna*) is differently influenced by visible and UV light. *Comp. Physiol. A* **183**, 709–717 (1998).
23. Leech, D. M. & Williamson, C. E. *In situ* exposure to ultraviolet radiation alters the depth distribution of *Daphnia*. *Limnol. Oceanogr.* **46**, 416–420 (2001).
24. Damkaer, D. M. in *The Role of Solar Ultraviolet Radiation in Marine Ecosystems* (ed. Calkins, J.) 701–706 (Plenum, New York, 1982).
25. Dodson, S. I. Predicting diel vertical migration of zooplankton. *Limnol. Oceanogr.* **35**, 1195–1200 (1990).
26. Smith, R. C. & Baker, K. S. Optical properties of the clearest natural waters (200–800 nm). *Appl. Optics* **20**, 177–184 (1981).
27. Vinyard, G. L. & O'Brien, W. J. Effects of light and turbidity on the reactive distance of bluegill (*Lepomis macrochirus*). *J. Fish. Res. Bd. Can.* **33**, 2845–2849 (1976).
28. Confer, J. L. *et al.* Visual predation by planktivores. *Oikos* **31**, 27–37 (1978).
29. Madronich, S., McKenzie, R. L., Caldwell, M. M. & Bjorn, L. Changes in ultraviolet radiation reaching the earth's surface. *Ambio* **24**, 143–152 (1995).

Acknowledgements

We thank L. J. Weider for providing arctic *D. pulex*. We also thank A. Agrawal, W. Gabriel, S. Giebler, A. F. Sell, B. Schierwater, T. Vines, C. E. Williamson, W. C. Kerfoot and L. J. Weider for comments and advice on earlier versions of the manuscript.

Correspondence and requests for materials should be addressed to R.T. (e-mail: ralph.tollrian@lrz.uni-muenchen.de).

Partitioning selection and complementarity in biodiversity experiments

Michel Loreau* & Andy Hector†

* Laboratoire d'Écologie, UMR 7625, École Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05, France

† NERC Centre for Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK

The impact of biodiversity loss on the functioning of ecosystems and their ability to provide ecological services has become a central issue in ecology. Several experiments have provided evidence that reduced species diversity may impair ecosystem processes such as plant biomass production^{1–5}. The interpretation of these experiments, however, has been controversial^{6–12} because two types of mechanism may operate in combination^{6,13–15}. In the 'selection effect', dominance by species with particular traits affects ecosystem processes. In the 'complementarity effect', resource partitioning or positive interactions lead to increased total resource use. Here we present a new approach to separate the two effects on the basis of an additive partitioning analogous to the

Price equation in evolutionary genetics^{16–19}. Applying this method to data from the pan-European BIODEPTH experiment⁴ reveals that the selection effect is zero on average and varies from negative to positive in different localities, depending on whether species with lower- or higher-than-average biomass dominate communities. In contrast, the complementarity effect is positive overall, supporting the hypothesis that plant diversity influences primary production in European grasslands through niche differentiation or facilitation.

Recent theoretical work has revealed that the observed responses of ecosystem processes to changes in species or functional-group diversity can be generated by a combination of different effects^{13–15}. These biodiversity effects can be grouped into two classes. First, there are those that arise from niche differentiation or facilitation between species, and that can increase the performance of communities above that expected from the performance of individual species. Distinguishing the effects of niche differentiation and facilitation may often be difficult in practice; therefore, we refer to these mechanisms collectively as 'complementarity'. One common form of complementarity in plant communities (which involves both resource partitioning and facilitation) arises between legumes, which have the ability to fix atmospheric nitrogen, and other plants, which have access only to soil nitrogen.

The second class of biodiversity effects gives rise to relationships between biodiversity and ecosystem functioning through selective processes, such as interspecific competition, which cause dominance (high relative abundance) of species with particular traits. For example, in one model of the 'sampling effect'^{6,13}, higher-diversity plant mixtures assembled at random from a pool of species have a higher chance of containing and becoming dominated by the

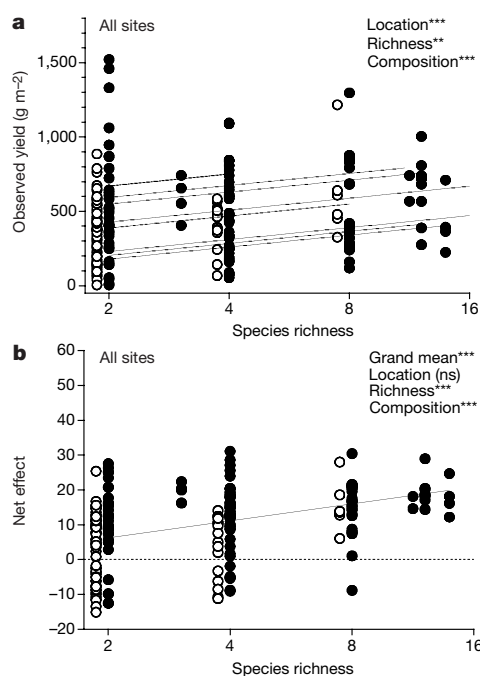


Figure 1 Observed yield Y_0 and net biodiversity effect ΔY as functions of species richness across all localities in mixtures of the BIODEPTH experiment. **a**, Observed yield; **b**, net biodiversity effect. Open circles are plots that do not contain any legume species; filled circles are plots that contain legumes. Lines are slopes from the multiple regression model using species richness on a $\log_2$ scale. Lines in **a** from highest elevation to lowest are: Germany, Silwood, Sheffield, Switzerland, Ireland, Greece, Sweden and Portugal. Values of the biodiversity effect (in g m^{-2}) are square-root transformed to meet the assumptions of analyses but preserve the original positive and negative signs. (Single asterisk, $P < 0.05$; double asterisk, $P < 0.01$; triple asterisk, $P < 0.001$; ns: non-significant.)

species that achieves the highest biomass when grown alone. The model assumes that the biomass of a mixture of species is equal to the monoculture biomass of the most productive of the component species.

Although both types of effect are potentially relevant, they have different implications and may apply to different circumstances¹⁵.

Box 1

Additive partition of biodiversity effects

We measure the net biodiversity effect, ΔY , by the difference between the observed yield of a mixture and its expected yield under the null hypothesis that there is no selection effect or complementarity effect. This expected value is the weighted (by the initial relative abundance of species in mixture) average of the monoculture yields for the component species. Positive selection occurs if species with higher-than-average monoculture yields dominate the mixtures. The selection effect is measured by the covariance between the monoculture yield of species and their change in relative yield in the mixture. Finally, a positive complementarity effect occurs if species yields in a mixture are on average higher than expected on the basis of the weighted average monoculture yield of the component species. These various effects can be related by additive partition as follows.

Define, for any mixture,

M_i = yield of species i in monoculture

Y_{O_i} = observed yield of species i in the mixture

$Y_O = \sum_i Y_{O_i}$ = total observed yield of the mixture

$RY_{E_i} =$ expected relative yield of species i in the mixture, which is simply its proportion seeded or planted

$RY_{O_i} = Y_{O_i}/M_i$ = observed relative yield of species i in the mixture

$Y_{E_i} = RY_{E_i}M_i$ = expected yield of species i in the mixture

$Y_E = \sum_i Y_{E_i}$ = total expected yield of the mixture

$\Delta Y = Y_O - Y_E$ = deviation from total expected yield in the mixture

$\Delta RY_i = RY_{O_i} - RY_{E_i}$ = deviation from expected relative yield of species i in the mixture

N = number of species in the mixture

It then follows that

$$\begin{aligned}\Delta Y &= Y_O - Y_E = \sum_i RY_{O_i}M_i - \sum_i RY_{E_i}M_i = \sum_i \Delta RY_i M_i \\ &= N \overline{\Delta RY} \overline{M} + N \text{cov}(\Delta RY, M)\end{aligned}$$

In this equation, $N \overline{\Delta RY} \overline{M}$ measures the complementarity effect, and $N \text{cov}(\Delta RY, M)$ measures the selection effect.

Note that this approach is a generalization of the relative yield total (RYT) and proportional deviation from expected value (D) approaches^{21,22}. Indeed, $N \overline{\Delta RY} = \text{RYT} - 1 = \overline{D}$ and $\Delta Y = D_T Y_E$. Thus, the above equation can be rewritten as $D_T Y_E = \overline{D} \overline{M} + N \text{cov}(\Delta RY, M)$ from which it is apparent that the net biodiversity effect is proportional to D_T , and the complementarity effect is proportional to $\overline{D}$.

As a consequence, our method has similar strengths and limitations to these approaches. Relative yield has traditionally been used in short-term plant competition experiments with substitutive designs in which total density is kept constant (as was the case in BIODEPTH). Results may then be dependent on the chosen density^{11,20}, although they seldom vary strongly in practice^{12,25}. Our approach, however, could be applied to other organisms and other experimental designs in which density is allowed to reach natural levels. As it is based on variation in relative yield (ΔRY), it could also be used to study how biodiversity effects change through time, by comparing observed values at each time with expected values calculated either from initial conditions or from the conditions of the previous time. □

Much of the controversy over recent experiments has revolved around the significance of the sampling effect and its contribution to the observed responses to experimental manipulations of biodiversity^{6–12}, prompting the need for methods to separate the sampling effect from complementarity^{20–22}. The sampling effect model, however, has restrictive assumptions which do not qualify it as a general alternative to complementarity. First, communities may not be dominated by a single species, and relative abundance and the magnitude of impact on ecosystem functioning may not be positively correlated. In particular, a negative selection effect can arise if communities are dominated by species with low—not high—values for a particular trait or process^{14,15,23}. Second, the sampling effect model actually combines two different processes: a sampling process and an extreme selection process that favours the single most productive species. The sampling process is shared by the selection and complementarity mechanisms: in both cases, communities that have more species also have a greater probability of containing the appropriate species (either single species with particular trait values or a combination of species with complementary traits). Thus selection, not sampling, really distinguishes the two classes of mechanism¹⁵. This strengthens the need for more general, conceptually clear methods that can identify their individual contributions in biodiversity experiments.

Our new methodological approach meets this need by providing an additive partitioning of two biodiversity effects: a 'selection effect' and a 'complementarity effect' (Box 1). The selection effect is based on Price's general theory of selection²⁴: selection occurs when changes in the relative yields of species in a mixture are non-randomly related to their traits (yields) in monoculture. Accordingly, selection is measured by a covariance function as in the Price equation of evolutionary genetics^{16–19}. The complementarity effect measures any change in the average relative yield in the mixture, whether positive (resulting from resource partitioning or facilitation) or negative (resulting from physical or chemical interference). The sum of these two effects is the net biodiversity effect; it measures the deviation of the mixture yield from its expected value on the basis of monoculture yields and the relative abundance of species in the mixtures. All three effects have the dimension of yield (where yield is a surrogate for any measurable ecosystem property), and an expected value of zero under the null hypothesis of no biodiversity effect. Consequently, all three can equally be positive or negative, and there is a potential for complementarity and selection effects to cancel each other, resulting in a zero net effect. Our additive partition unifies and relates in a single equation previous measures based on the relative yield total and proportional deviation from expected value approaches (Box 1). Another advantage over previous methods is that our additive partition provides absolute measures of biodiversity effects, thereby allowing quantitative comparison of their respective contributions.

We applied this methodology to patterns of aboveground biomass production obtained from the BIODEPTH project, a network of field experiments that examined the functioning of European grassland ecosystems in relation to the direct manipulation of plant diversity⁴. As our approach requires a comparison between the performances of species in mixture and in monoculture, we restricted its application to the subset of experimental mixture plots that contained species for which monoculture yields were available. The overall pattern for this subset was similar to that for the whole experiment⁴: a log-linear increase in aboveground biomass with species richness and a significant locality effect, but no significant interaction between species richness and locality (Fig. 1a). The net biodiversity effect was positive (the grand mean was significantly different from zero) and increased significantly with species richness beyond two species (Fig. 1b; Table 1). Species composition always had a significant effect.

The two components of this net effect, selection and complementarity, had strikingly different patterns. The selection effect

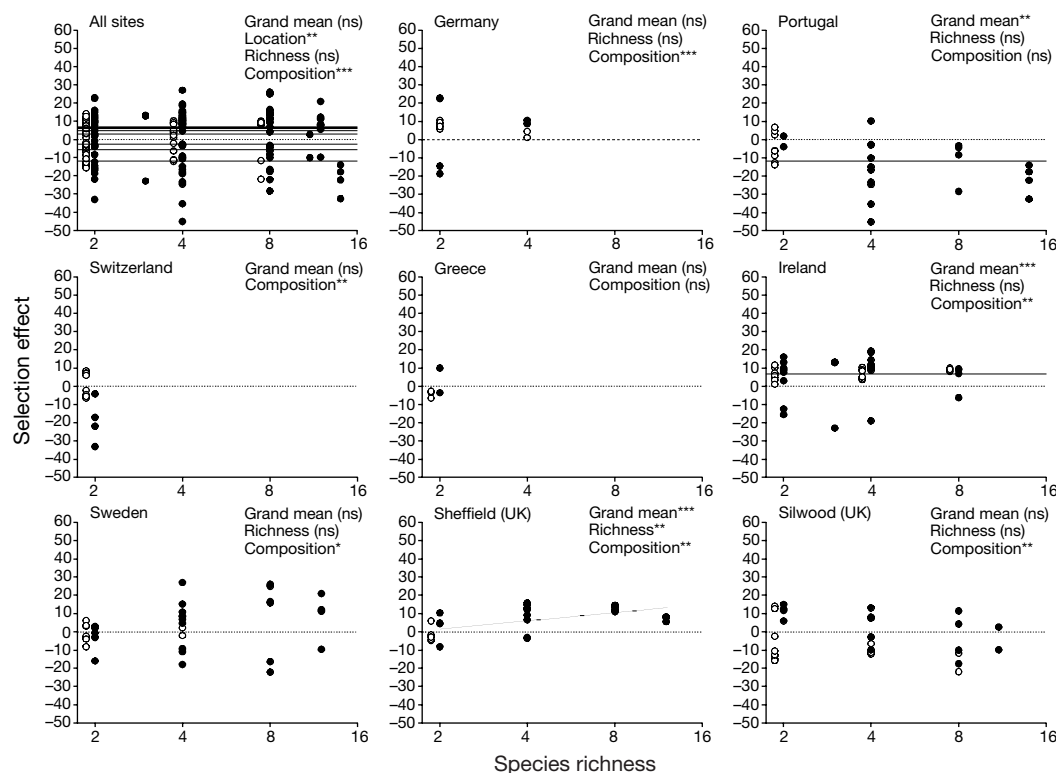


Figure 2 Selection effect, $N \text{ cov}(\Delta RY, M)$, as a function of species richness across all localities and at individual localities. Lines from highest elevation to lowest are: Ireland, Germany, Sheffield, Greece, Sweden, Silwood, Switzerland and Portugal. Values are

square-root transformed to meet the assumptions of analyses but preserve the original positive and negative signs. Symbols as in Fig. 1.

was variable, ranging from significantly positive averages in two localities (Ireland and Sheffield) to a significantly negative average in one locality (Portugal) (Fig. 2). Overall, across all sites, these variations cancelled out, so that the grand mean was not significantly different from zero, and selection was unaffected by species richness. The only factors that influenced selection significantly were locality and species composition (Fig. 2; Table 1).

In contrast, the average complementarity effect was significantly positive in four localities (Portugal, Ireland, Sheffield and Silwood) and in the overall analysis across site (Fig. 3; Table 1). Furthermore, complementarity increased significantly with the species richness of mixtures in two localities (Sheffield and Portugal) and across all localities. There were again significant locality and composition effects, but no significant locality-by-species richness interaction (Table 1). Thus, the best statistical model for the overall across-site analysis was a log-linear relationship between complementarity and species richness with identical slopes but different heights in the

various localities (Fig. 3).

One particular form of complementarity occurs between nitrogen-fixing legumes and other plants. Might the fertilization effect of legumes be sufficient to explain the positive complementarity effect found in this experiment¹¹? The presence of legumes in the mixtures did have important impacts on their performance; in general it tended to increase yields as well as the net and complementarity effects, and to generate more extreme selection effects, whether positive or negative (Figs 1–3). Although the experiment was not designed to fully separate the effects of legumes or functional-group diversity from that of species diversity, our method can also be applied to examine this issue. When the presence of legumes was included as an additional factor in our across-site analyses, species richness—in addition to locality, presence of legumes and species composition—retained a significant log-linear effect on complementarity, whether legumes were introduced after ($F_{(1,67)} = 11.36$, $P < 0.01$) or before ($F_{(1,67)} = 4.74$, $P < 0.05$)

Table 1 Summary of the analysis of biodiversity effects for mixtures of the BIODEPTH experiment

	d.f.	Net effect, ΔY			Selection effect, $N \text{ cov}(\Delta RY, M)$			Complementarity effect, $N \Delta RY M$		
		MS	F	P	MS	F	P	MS	F	P
Grand mean	1	20,000.65	129.30	2.63×10^{-18}	257.78	1.44	0.2334	15,830.07	87.74	1.87×10^{-14}
Locality	7	224.09	1.45	0.1980	1,267.88	7.09	0.0094	606.99	3.36	0.0034
Species richness ($\log_2$)	1	3,210.81	20.76	1.87×10^{-5}	23.35	0.13	0.7188	1,879.89	10.42	0.0018
Deviation	5	76.61	0.50	0.7789	133.54	0.75	0.5908	177.04	0.98	0.4346
Locality $\times$ richness	5	203.89	1.32	0.2650	340.27	1.90	0.1031	110.68	0.61	0.6898
Composition	79	154.68	4.41	1.14×10^{-12}	178.79	3.13	2.59×10^{-8}	180.42	1.99	0.0005
Residual	107	35.06			57.13			90.69		
Total	205									

All terms were tested against the average variation between replicate mixtures—the Composition term—and composition effects against the overall residual (mixture-by-block interaction). The first line reports a test of the grand mean effects across all mixtures versus the null expectation of zero and subsequent lines test for effects of location and diversity. Data were square-root transformed while preserving the original positive or negative signs to meet the assumptions of analyses. d.f., degrees of freedom; MS, mean squares.

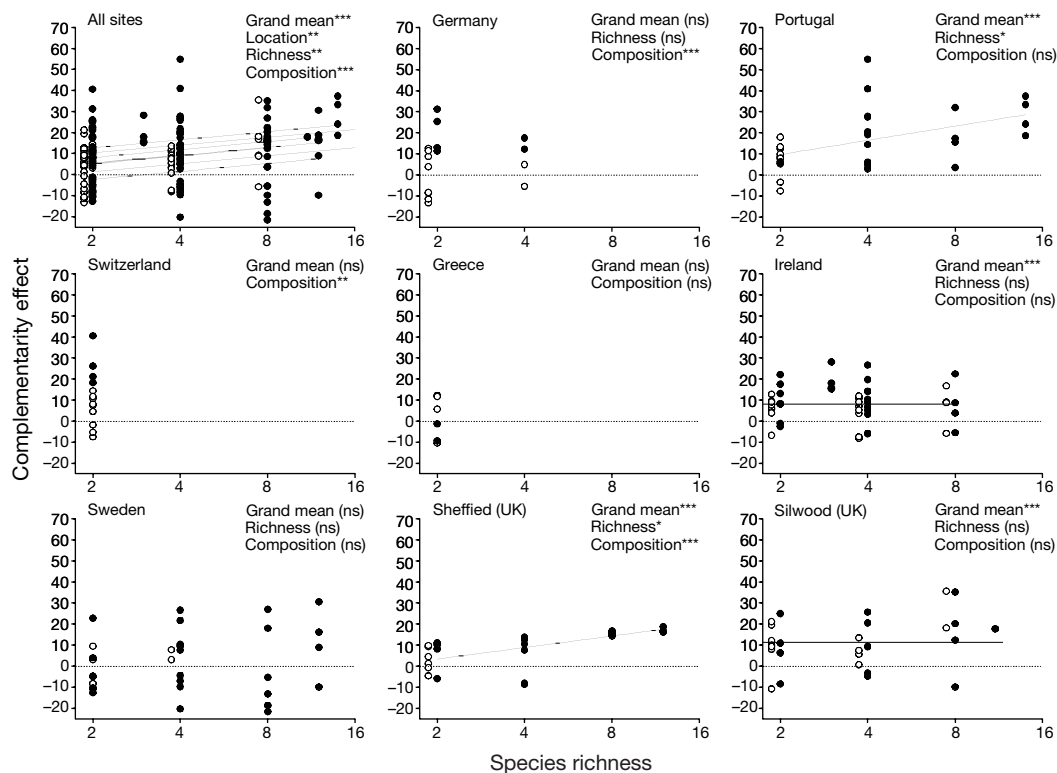


Figure 3 Complementarity effect, $N \Delta RY M$, as a function of species richness across all localities and at individual localities. Lines from highest elevation to lowest are: Portugal, Switzerland, Silwood, Germany, Sheffield, Ireland, Greece and Sweden. Values are

square-root transformed to meet the assumptions of analyses but preserve the original positive and negative signs. Symbols as in Fig. 1.

species richness. The latter test was very conservative as the presence of legumes was embedded within species richness in the experiment, so that part of the species richness effect was absorbed into the legume effect when it was fitted first in the analysis. Thus, the increased complementarity in species-rich mixtures involved not only complementarity between legumes and other plant types, but also complementarity between species within each of these groups.

Our analysis confirms theoretical predictions that selection effects should be more variable and less predictable than complementarity^{14,15}. In the BIODEPTH experiment, selection favoured species with either high or low monoculture yield depending on locality, and did not contribute to aboveground biomass patterns across localities. Therefore the selection effect can be rejected as the sole mechanism explaining the results from this experiment. In contrast, although complementarity was also variable, it was consistently positive overall and presented the same log-linear dependency on species richness as did the raw and net biodiversity effects. Furthermore, this dependence involved both complementarity between nitrogen-fixing legumes and other species, and complementarity between species within each group. Therefore, our analysis supports the hypothesis that plant diversity influences primary production in European grasslands at least partly through some combination of niche differentiation and facilitation between species⁴.

Our methodology provides a powerful means to *a posteriori* partition the selection and complementarity components of biodiversity effects in biodiversity–ecosystem functioning experiments. Our measure of complementarity, however, is linearly related to the relative yield total and, therefore, has similar strengths and limitations (Box 1). It cannot replace direct experimental investigations into the mechanisms at work in responses to biodiversity changes at the ecosystem level, which are now critical to further progress in this area. □

Methods

Experimental data

The BIODEPTH experiment simulated the loss of plant species in grassland ecosystems by removing the existing vegetation and seed bank and re-establishing plant communities from seed using standardized protocols at two localities in the UK (Sheffield and Silwood) and at single localities in Germany, Ireland, Greece, Portugal, Sweden and Switzerland. At each site, five levels of species richness were established, ranging from monocultures to higher diversity assemblages that approximately matched background levels of diversity in comparable unmanipulated semi-natural grasslands. To replicate plant diversity, each level of species richness was represented by several different plant assemblages at each site. Constrained random selection from the local pool of grassland species was used to form assemblages where all polycultures contained at least one grass. To investigate the effects of species composition, each assemblage was replicated in a minimum of two plots at each site including monocultures of many of the species involved. In total, the experiment comprised 480 plots, in which various ecosystem properties were measured. More details on the experimental design can be found elsewhere^{4,12}. For our analysis, we used data on aboveground plant biomass production measured two years after establishment of the experimental plots, in the subset of polyculture plots for which all species were also grown in monoculture at the same site. This subset comprised 205 of the 308 polyculture plots.

Statistical analyses

Data were analysed by analysis of variance and linear regression. Initial (sown) species richness was used in the analysis because it defined treatments in the experimental design; realised species richness closely matched sown species richness and produced similar results^{4,12}. Values of the selection, complementarity and net biodiversity effects were transformed to meet assumptions of tests by taking the square root while preserving the original sign (alternative transformations produced similar results). We tested two main hypotheses. First, we tested grand mean values across all polycultures versus zero to see whether they differed significantly from the weighted average of the monoculture yields. Second, we performed analysis of variance, partitioning the relationship into a linear regression, testing for positive or negative relationships with species richness above two species, and testing the deviation from this relationship. Details of similar analyses of the full set of 480 plots are available elsewhere^{4,12}.

Received 20 December; accepted 13 March 2001.

1. Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
2. Tilman, D. *et al.* The influence of functional diversity and composition on ecosystem processes. *Science* **277**, 1300–1302 (1997).

3. van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
4. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
5. Schläpfer, F. & Schmid, B. Ecosystem effects of biodiversity: a classification of hypotheses and exploration of empirical results. *Ecol. Appl.* **9**, 893–912 (1999).
6. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
7. Aarssen, L. W. High productivity in grassland ecosystems: effected by species diversity or productive species? *Oikos* **80**, 183–184 (1997).
8. Tilman, D. Distinguishing the effects of species diversity and species composition. *Oikos* **80**, 185 (1997).
9. Wardle, D. A. Is 'sampling effect' a problem for experiments investigating biodiversity–ecosystem function relationships? *Oikos* **87**, 403–407 (1999).
10. van der Heijden, M. G. A. *et al.* 'Sampling effect', a problem in biodiversity manipulation? A reply to David A. Wardle. *Oikos* **87**, 408–410 (1999).
11. Huston, M. A. *et al.* No consistent effect of plant diversity on productivity. *Science* **289**, 1255 (2000).
12. Hector, A. *et al.* No consistent effect of plant diversity on productivity: Response. *Science* **289**, 1255 (2000).
13. Tilman, D., Lehman, C. & Thompson, K. Plant diversity and ecosystem productivity: theoretical considerations. *Proc. Natl Acad. Sci. USA* **94**, 1857–1861 (1997).
14. Loreau, M. Biodiversity and ecosystem functioning: a mechanistic model. *Proc. Natl Acad. Sci. USA* **95**, 5632–5636 (1998).
15. Loreau, M. Biodiversity and ecosystem functioning: recent theoretical advances. *Oikos* **91**, 3–17 (2000).
16. Price, G. R. Selection and covariance. *Nature* **227**, 520–521 (1970).
17. Price, G. R. Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**, 485–490 (1972).
18. Frank, S. A. George Price's contributions to evolutionary genetics. *J. Theor. Biol.* **175**, 373–388 (1995).
19. Frank, S. A. The Price Equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**, 1712–1729 (1997).
20. Garnier, E., Navas, M.-L., Austin, M. P., Lilley, J. M. & Gifford, R. M. A problem for biodiversity–productivity studies: how to compare the productivity of multispecific plant mixtures to that of monocultures? *Acta Oecol.* **18**, 657–670 (1997).
21. Hector, A. The effect of diversity on productivity: detecting the role of species complementarity. *Oikos* **82**, 597–599 (1998).
22. Loreau, M. Separating sampling and other effects in biodiversity experiments. *Oikos* **82**, 600–602 (1998).
23. Troumbis, A. Y., Dimitrakopoulos, P. G., Siamantziouras, A.-S. D. & Memtsas, D. Hidden diversity and productivity patterns in mixed Mediterranean grasslands. *Oikos* **90**, 549–559 (2000).
24. Price, G. R. The nature of selection. *J. Theor. Biol.* **175**, 389–396 (1995).
25. Jolliffe, P. A. The replacement series. *J. Ecol.* **88**, 371–385 (2000).

Acknowledgements

We thank S. Yachi, B. Schmid and D. Deutschman for suggestions and all members of BIODDEPTH for their input to the project. This work was supported by the CNRS GDR 1936 DIV-ECO and the CNRS PICS *Dynamics of biodiversity: from species interactions to ecosystem functioning*.

Correspondence and requests for materials should be addressed to M.L. (e-mail: Loreau@ens.fr).

Cryptic evolution in a wild bird population

J. Merilä*, L. E. B. Kruuk† & B. C. Sheldon‡§

* Department of Population Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18d, SE-752 36 Uppsala, Sweden

† Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

‡ Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

§ Department of Animal Ecology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18d, SE-752 36 Uppsala, Sweden

Microevolution is expected to be commonplace, yet there are few thoroughly documented cases of microevolution in wild populations^{1,2}. In contrast, it is often observed that apparently heritable traits under strong and consistent directional selection fail to show the expected evolutionary response^{3,4}. One explanation proposed for this paradox is that a genetic response to selection may be masked by opposing changes in the environment^{5,6}. We used data from a 20-year study of collared

flycatchers (*Ficedula albicollis*) to explore selection on, and evolution of, a heritable trait: relative body weight at fledging ('condition'). Despite consistent positive directional selection, on both the phenotypic and the additive genetic component (breeding values, estimated from an animal model) of condition, the mean phenotypic value of this trait in the population has declined, rather than increased, over time. Here we show that, despite this decline, the mean breeding value for condition has increased over time. The mismatch between response to selection at the levels of genotype and phenotype can be explained by environmental deterioration, concealing underlying evolution. This form of cryptic evolution may be common in natural environments.

If selection acts consistently on a heritable trait in a population, it should, all else being equal, induce a permanent change in the distribution of that trait⁷. The frequent lack of expected evolutionary change in heritable traits under directional selection in the wild has therefore puzzled evolutionary biologists for some time. Explanations proposed to account for this paradox include: inflated estimates of heritability owing to environmental covariance between relatives⁷, spatially and temporally varying selection pressures⁸, negative genetic correlations between different components of fitness⁸, and selection restricted to the environmental component of the phenotype^{3,4}. Another possibility is that a genetic response to selection does in fact occur, but is masked by opposing changes in the environment^{5,6}. However, to date, these alternatives have been subjected to very little empirical scrutiny⁸.

In many passerine bird species, relative body mass (the condition index) is an important predictor of the survival of fledglings: relatively heavier nestlings are more likely to survive to become breeding adults^{9–11}. This is also true for juvenile survival in other taxa, such as reptiles¹² and mammals¹³. In the collared flycatcher, quantitative genetic analyses using traditional methods suggest a

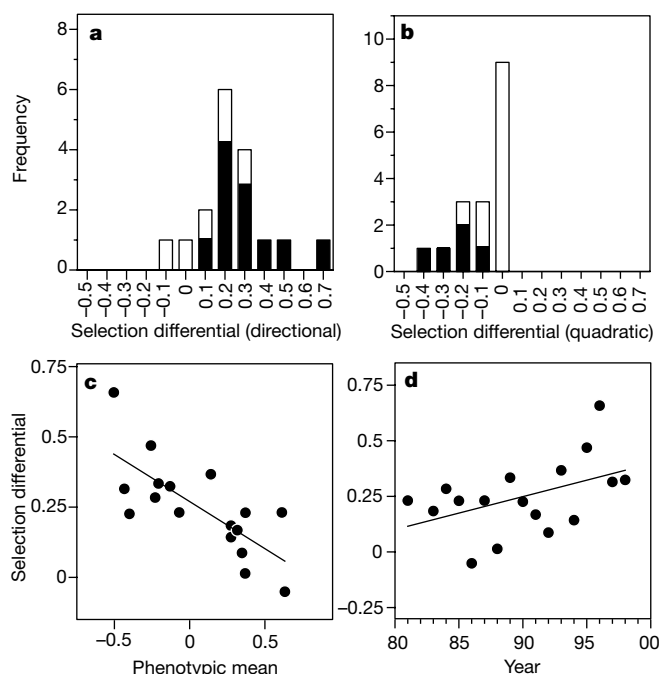


Figure 1 Patterns of natural selection on condition index of nestling collared flycatchers from 1981 to 1998. **a**, Distribution of directional standardized selection coefficients observed in each year. **b**, Distribution of quadratic standardized selection coefficients, with negative values indicating stabilizing selection. **c**, Relationship between mean condition index and intensity of directional survival selection across different study years. **d**, Intensity of directional survival selection on body condition index as a function of time. In **a** and **b**, the shaded proportions of bars indicate selection coefficients significantly ($P < 0.05$) different from zero.

significant additive genetic component to variation in the body condition index¹⁴. Mixed model analysis of variance using data for 17,717 offspring in 3,836 breeding attempts, from a long-term study of this species on the island of Gotland, Sweden, confirms this finding, revealing a narrow-sense heritability of 0.30 (standard error, s.e. = 0.03; ref. 15). Analyses of survival selection show that there is significant positive directional selection on condition, such that the survivors are, on average, 0.23 (s.e. = 0.02; $P < 0.001$) standard deviations above the population mean (Fig. 1a; ref. 15). In some years (6 out of 17), there is also significant stabilizing selection acting on condition, but this is both weaker and less consistent than the directional selection (Fig. 1b). Given that the condition index is heritable and under positive directional selection, we would expect the mean in this population to be evolving towards higher values. However, in contrast to this expectation, the mean condition in the population decreased significantly between 1981 and 1999 (linear regression of annual means: $b = -0.036$, s.e. = 0.015, $t_{15} = 2.35$, $P = 0.032$; Fig. 2a; generalized linear mixed model (GLMM) of individual values, controlling for family structure and non-independence of observations from the same year: $b = -0.035$, s.e. = 0.003, $t_{16,057} = 10.25$, $P < 0.001$).

One plausible explanation for the lack of response to selection is that it is predominantly the non-heritable, or environmental, component of the condition index that determines survival^{3,4}, such that natural selection does not act on the genetic component of variation in condition. A role for environmental variation in natural selection is suggested by the observation that selection tends to be strongest in years when the mean condition is lowest (linear regression: $b = -0.315$, s.e. = 0.079, $t_{15} = 4.24$; $P < 0.001$; Fig. 1c). However, a direct test is to calculate selection on estimated breeding values, or the expected effect of the genes that an individual passes on to its offspring, which can be derived from pedigree information⁷. This analysis shows that selection acts directly on breeding values (standardized selection differential, $S = 0.14$, s.e. = 0.02, $P < 0.001$), not only on environmental deviations¹⁵. Furthermore, there is no evidence that a response to selection on condition would be constrained by negative genetic correlations with other fitness components. Both individual life-

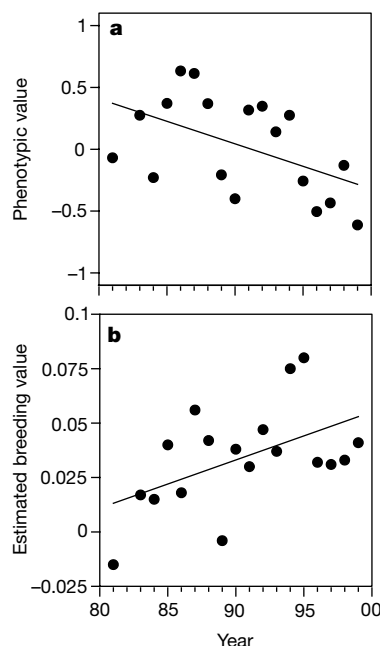


Figure 2 Changes in condition index over time in the collared flycatcher population from 1981 to 1999. **a**, Mean phenotypic value of condition index. **b**, Mean estimated breeding value of condition index. See text for statistical tests.

span (LSP) and lifetime reproductive success (LRS) are positively correlated with the breeding values for condition (GLMM: LSP, $\chi^2_{(1)} = 9.88$, $P = 0.002$; LRS, $\chi^2_{(1)} = 21.42$, $P < 0.001$; Fig. 3), and the genetic correlations between condition and LSP and between condition and LRS were +0.009 (s.e. = 0.151) and -0.066 (s.e. = 0.265), respectively: neither of these correlations was significantly different from zero.

An alternative explanation for the apparent paradox—the absence of any evolutionary change despite significant directional selection on a heritable trait—lies in the possibility of change in environmental conditions over the study period^{5,6}. This explanation runs parallel to that suggested to account for the apparent lack of genetic differentiation across environmental gradients when such differentiation is expected—a phenomenon referred to as countergradient variation¹⁶ (Fig. 4a). Countergradient variation is defined as a negative covariance between the environmental and genetic influences on a given trait across some environmental gradient, and it can effectively conceal genetic differentiation when the environmental influence is sufficiently strong¹⁶. In the context of the current study the countergradient hypothesis would predict that, at the genotypic level, there should be a positive correlation between year of the study and condition index. Using estimated breeding values we found that the mean estimated breeding value had indeed increased over the course of the study (linear regression of annual means: $b = 0.0022$, s.e. = 0.0009, $t_{15} = 2.38$, $P = 0.030$; GLMM of individual values: $b = 0.0023$, s.e. = 0.0008, $t_{16,057} = 2.70$, $P = 0.007$; Fig. 2b). Hence, despite the negative trend at the phenotypic level (Fig. 2a), at the level of the genotype the population mean condition index has increased over time.

The estimated microevolutionary change has presumably been concealed by an increasingly negative influence of environmental

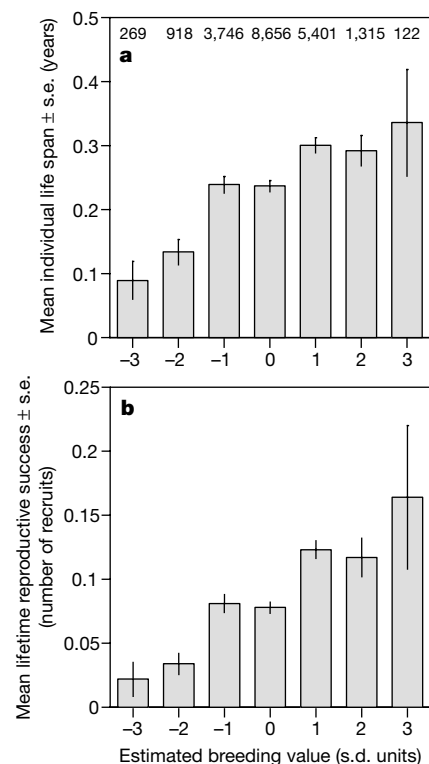


Figure 3 Associations between key life history traits and estimated breeding values (EBVs) for condition. **a**, Individual life-span (in years); and **b**, mean lifetime reproductive success (LRS; number of recruits produced during an individual's lifetime) as a function of EBV. EBVs have been truncated to bins of one standard deviation (s.d.), each containing individuals with EBVs ± 0.5 s.d. units around the bin-centre. Numbers at the top of **a** refer to the number of individuals on which the means are based. See text for statistical tests.

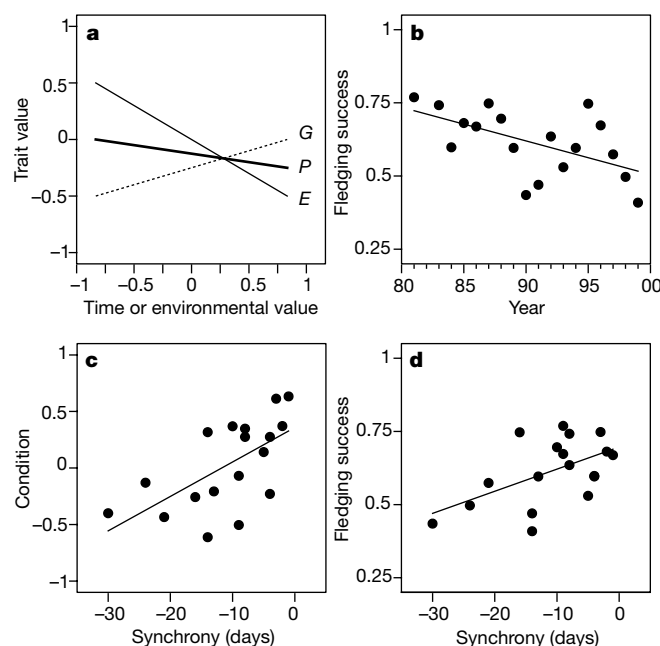


Figure 4 Environmental deterioration over time. **a**, A simple graphical model illustrating how directional change in environmental conditions over time (x-axis) may conceal genetic evolution by inducing a change in the opposite direction at the phenotypic level. *P*, phenotypic mean; *G*, genotypic mean; *E*, effect of environment on trait value. Note that the x-axis could also represent some environmental gradient and that $P = G + E$ (ref. 16). **b**, Average fledging success (proportion of laid eggs producing individuals

leaving the nest) of collared flycatchers as a function of time. **c**, Average condition index and **d**, average fledging success of collared flycatchers as a function of synchrony between caterpillar emergence and oak bud-burst dates in different study years. Synchrony index is defined as the number of days elapsed between caterpillar emergence and oak bud-burst in the Netherlands¹⁷. See text for details.

conditions on the condition index (Fig. 4a), which has caused the phenotypic decline. The intensity of selection on the condition has increased with time (linear regression: $b = 0.015$, s.e. = 0.007, $t_{17} = 1.99$, $P = 0.06$; Fig. 1d), and fledging success has decreased with time (Fig. 4b; linear regression: $b = -0.011$, s.e. = 0.003, $t_{18} = 3.19$, $P < 0.005$). Both relationships are indicative of environmental deterioration. A plausible agent explaining this deterioration is the large-scale climatic trend that has reduced the caterpillar food supply—the main food of growing nestlings—over the last few decades¹⁷. Increased spring temperatures have led to increasingly poor synchronization between the hatching date of caterpillars and the date of bud-burst of the oak trees on which they feed. Estimates of the lag between caterpillar hatching date and oak bud-burst date from a Dutch study¹⁷ were positively correlated with the annual mean condition index (coefficient of correlation $r = 0.613$, test statistic $z_{17} = 2.76$, $P = 0.0057$; Fig. 4c), tarsus length ($r = 0.621$, $z_{17} = 2.89$, $P = 0.0021$) and fledging success ($r = 0.531$, $z_{17} = 2.82$, $P = 0.0049$; Fig. 4d) of collared flycatchers on Gotland. Because the degree of synchrony between caterpillar emergence and bud-burst dates is driven by large-scale climatic phenomena¹⁷, such correlations can be expected to occur across continental scales. Increased intra- and interspecific competition, both of which lower condition in this^{18,19} and other bird populations²⁰ are other potential, but perhaps less likely, explanations for the decline in reproductive success and condition observed in this study.

In conclusion, in accordance with data from studies of spatial genetic differentiation¹⁶, our results show that microevolutionary transitions at the genotypic level need not necessarily be manifested at the phenotypic level¹⁶, and that an apparent lack of evolutionary response in a heritable trait subject to directional natural selection can be understood in terms of the environment masking genotypic evolution, rather than selection on environmental deviations only³. Alternative explanations for the lack of selection response (that is, biased heritability estimates, negative genetic correlations between the focal trait and other components of fitness, reversed direction of

selection on later life stages) could be excluded on the basis of detailed analysis using new methods. To this end, our results concur with the view^{21,22} that many evolutionary transitions may consist of changes not visible at the level of the phenotype. □

Methods

The data

The material for this study was collected between 1980 and 1999 from a nest-box-breeding collared flycatcher population inhabiting the island of Gotland, off the Swedish east coast in the Baltic sea. All breeding attempts were monitored from the date of egg-laying until all nestlings had fledged. When 12 days old, nestlings were measured for tarsus length with digital calipers (to nearest 0.1 mm), weighed with a Pesola spring balance (to nearest 0.1 g) and marked with individually numbered aluminium rings. At the same time, their parents were captured and their identity was checked (see refs 14 and 15 for more information on collection procedures). Condition index was estimated as the residuals from a linear regression of body mass at fledging on tarsus length (see refs 10 and 14 for details and analysis of linearity of this relationship). Data were available for 4,888 breeding attempts involving 23,336 individuals. Because the sexes do not differ in growth patterns or condition at fledging²³, data on both sexes were analysed together (with one exception: see below). Breeding attempts subject to manipulative experiments were excluded from the analyses.

Quantitative genetic analyses

Heritability of condition and individual breeding values were estimated through a mixed model restricted maximum likelihood (REML) estimation procedure using the software packages VCE²⁴ and PEST²⁵. An individual 'animal model' was fitted, in which an individual's phenotypic value of condition was broken down into components of additive genetic value and other random and fixed effects^{7,26}. The area of the study site and the year were included as random effects to account for spatial and temporal heterogeneity in environmental effects on phenotype. Nest-box identity was also fitted as a random effect to account for further common-environment effects specific to the individual brood; this term will incorporate any non-genetic maternal effects. The only fixed effect in the model was a population mean. The narrow-sense heritability (h^2) was estimated as the ratio of the additive genetic variance (V_A) to the total phenotypic variance (V_P): $h^2 = V_A/V_P$. Best linear unbiased predictors (BLUP) of individual breeding values were quantified from pedigree information using REML estimates of variance components, with the software package PEST²⁵. BLUP estimates of breeding values (EBVs) are unbiased even in populations under selection, or exhibiting assortative mating, and estimates from different generations will reflect changes in additive genetic effects resulting from selection, genetic drift or inbreeding⁷. With an annual breeding population size in excess of 1,500 pairs, the occurrence of genetic drift in this population is unlikely. Calculation of

inbreeding coefficients (IBCs) revealed that only 0.5% of pairings resulted in individuals with an IBC greater than 0, and there was no evidence for between-year heterogeneity in IBCs (Kruskal–Wallis $H_{18} = 15.15$, $P = 0.65$; estimated using Pedigree Viewer, available from <http://www.personal.une.edu.au/~bkinghor/pedigree.htm>). Changes in EBVs across generations can therefore be taken as evidence of a response to selection, or 'genetic trend'^{27,28}. Genetic correlations were estimated from a multivariate animal model analysis of fledgling condition, life-span (LSP; in years) and lifetime reproductive success (LRS; defined as the number of offspring recruited into the breeding population). Genetic correlation analyses were necessarily restricted to individuals surviving to adulthood; area and year were included as random effects, and sex (known for individuals recaptured as adults) as a fixed effect.

Selection analyses

Estimates of survival selection on phenotypic and estimated breeding values of condition index were based on recapture data under the assumption that nestlings not returning to the study area in subsequent years had died. As many of the individuals recruit to the population at the age of two years, the survival analyses were restricted to the period of 1981–1998. Standardized directional (S) and quadratic (c^2) selection differentials were estimated by linear regression of relative fitness on standardized (zero mean, unit variance) phenotypic or breeding values of the condition index using standard methods²⁹. Statistical significance of the selection differentials was estimated with logistic regression²⁹. Associations between individual LSP or LRS and EBVs for condition were tested using GLMMs with negative binomial error structure, using the procedure IRREML in Genstat³⁰. Nest of origin, year and area were included as random effects in the model, to account for repeated measures. The significance of the fixed effect of condition breeding value as a predictor of LSP or LRS was assessed by the Wald statistic, distributed as $\chi^2_{(1)}$ (ref. 30).

Received 26 February; accepted 14 May 2001.

- Grant, P. R. & Grant, B. R. Predicting microevolutionary responses to directional selection on heritable variation. *Evolution* **49**, 241–251 (1995).
- Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. Evaluation of the rate of evolution in a natural population of guppies. *Science* **275**, 1934–1937 (1997).
- Price, T., Kirkpatrick, M. & Arnold, S. J. Directional selection and the evolution of breeding date in birds. *Science* **240**, 798–799 (1988).
- Alatalo, R. V., Gustafsson, L. & Lundberg, A. Phenotypic selection on heritable size traits: environmental variance and genetic response. *Am. Nat.* **135**, 464–471 (1990).
- Cooke, F., Taylor, P. D., Frances, C. M. & Rockwell, R. F. Directional selection and clutch size in birds. *Am. Nat.* **136**, 261–267 (1990).
- Frank, S. A. & Slatkin, M. Fisher's fundamental theorem of natural selection. *Trends Ecol. Evol.* **7**, 92–95 (1992).
- Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, Sunderland, Massachusetts, 1998).
- Roff, D. A. *Evolutionary Quantitative Genetics* (Chapman & Hall, New York, 1997).
- Hochachka, W. & Smith, J. N. Determinants and consequences of nestling condition in song sparrows. *J. Anim. Ecol.* **60**, 995–1008 (1991).
- Lindén, M., Gustafsson, L. & Pärt, T. Selection of fledgling mass in the collared flycatcher and the great tit. *Ecology* **73**, 336–343 (1992).
- Both, C., Visser, M. E. & Verboven, N. Density-dependent recruitment rates in great tits: the importance of being heavier. *Proc. R. Soc. Lond. B* **266**, 465–469 (1999).
- Sorci, G. & Clobert, J. Natural selection on hatching body size and mass in two environments in the common lizard (*Lacerta vivipara*). *Evol. Ecol. Res.* **1**, 303–316 (1999).
- Boltnov, A. I., York, A. E. & Antonelis, G. A. Northern fur seal young: interrelationships among birth size, growth, and survival. *Can. J. Zool.* **76**, 843–854 (1998).
- Merilä, J. Genetic variation in offspring condition—an experiment. *Funct. Ecol.* **10**, 465–474 (1996).
- Merilä, J., Kruuk, L. E. B. & Sheldon, B. C. Natural selection on the genetical component of body condition in a wild bird population. *J. Evol. Biol.* (submitted).
- Conover, D. O. & Schultz, E. T. Phenotypic similarity and the evolutionary significance of counter-gradient variation. *Trends Ecol. Evol.* **10**, 248–252 (1995).
- Visser, M. E. & Holleman, L. J. M. Warmer springs disrupt the synchrony of oak and winter moth phenology. *Proc. R. Soc. Lond. B* **268**, 289–294 (2001).
- Gustafsson, L. Inter- and intraspecific competition for nest holes in a population of the collared flycatcher *Ficedula albicollis*. *Ibis* **130**, 11–16 (1988).
- Doligez, B., Danchin, E., Clobert, J. & Gustafsson, L. The use of conspecific reproductive success for breeding habitat selection in a non-colonial, hole-nesting species, the collared flycatcher. *J. Anim. Ecol.* **68**, 1193–1206 (2000).
- Alatalo, R. V. & Lundberg, A. Density-dependence in breeding success of the pied flycatcher (*Ficedula hypoleuca*). *J. Anim. Ecol.* **53**, 969–978 (1984).
- Lewontin, R. C. Adaptation. *Sci. Am.* **239**, 212–230 (1978).
- Gilchrist, A. S. & Partridge, L. A comparison of the genetic basis of wing size divergence in three parallel body size clines of *Drosophila melanogaster*. *Genetics* **153**, 1775–1787 (1999).
- Sheldon, B. C., Merilä, J., Lindgren, G. & Ellegren, H. Gender and environmental sensitivity in nestling collared flycatchers. *Ecology* **79**, 1939–1948 (1998).
- Groeneveld, E. REML VCE, a Multivariate Multi-Model Restricted Maximum Likelihood (Co) Variance Component Estimation Package, Version 3.2 User's Guide. (Institute of Animal Husbandry and Animal Behaviour, Federal Research Center of Agriculture (FAL), Mariensee, Germany, 1995). (<http://www.tzv.fal.de/institut/genetik/pub/eg/vce/manual/manual.html>).
- Groeneveld, E., Kovac, M., Wang, T. L. & Fernando, R. L. Computing algorithms in a general purpose BLUP package for multivariate prediction and estimation. *Arch. Anim. Breed.* **15**, 399–412 (1992).
- Meyer, K. Restricted maximum-likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet. Selection Evol.* **21**, 317–340 (1989).
- Blair, H. T. & Pollak, E. J. Estimation of genetic trend in a selected population with and without the use of a control population. *J. Anim. Sci.* **58**, 878–886 (1984).

- Southwood, O. I. & Kennedy, B. W. Genetic and environmental trends for litter size in swine. *J. Anim. Sci.* **69**, 3177–3182 (1991).
- Merilä, J., Sheldon, B. C. & Ellegren, H. Antagonistic natural selection revealed by molecular sex identification of nestling collared flycatchers. *Mol. Ecol.* **6**, 1167–1175 (1997).
- Genstat, 1998. Genstat 5, Release 4.1 (Lawes Agricultural Trust, IACR, Rothamsted, 1998).

Acknowledgements

We thank I. P. F. Owens, A. J. van Noordwijk, B. Walsh and D. A. Roff for comments on the manuscript, M. Visser for data on caterpillars and oaks, and the numerous people who have helped in collecting the data in the course of the study, in particular L. Gustafsson. Our research was supported by the Swedish Natural Science Research Council, the Nordic Academy for the Advanced Study (J.M.) and by Royal Society University Research Fellowships to B.C.S. and L.E.B.K.

Correspondence and requests for materials should be addressed to J.M.
(e-mail: juha.merila@ebc.uu.se).

Cortical remodelling induced by activity of ventral tegmental dopamine neurons

Shaowen Bao, Vincent T. Chan & Michael M. Merzenich

Keck Center for Integrative Neuroscience, University of California, San Francisco, California 94143, USA

Representations of sensory stimuli in the cerebral cortex can undergo progressive remodelling according to the behavioural importance of the stimuli^{1,2}. The cortex receives widespread projections from dopamine neurons in the ventral tegmental area (VTA)^{3–5}, which are activated by new stimuli or unpredicted rewards^{6,7}, and are believed to provide a reinforcement signal for such learning-related cortical reorganization⁸. In the primary auditory cortex (AI) dopamine release has been observed during auditory learning that remodels the sound-frequency representations^{9,10}. Furthermore, dopamine modulates long-term potentiation^{11,12}, a putative cellular mechanism underlying plasticity¹³. Here we show that stimulating the VTA together with an auditory stimulus of a particular tone increases the cortical area and selectivity of the neural responses to that sound stimulus in AI. Conversely, the AI representations of nearby sound frequencies are selectively decreased. Strong, sharply tuned responses to the paired tones also emerge in a second cortical area, whereas the same stimuli evoke only poor or non-selective responses in this second cortical field in naïve animals. In addition, we found that strong long-range coherence of neuronal discharge emerges between AI and this secondary auditory cortical area.

In our first experiment seven rats underwent twenty 2-h daily sessions in which 9-kHz pulsed tones were paired with electrical microstimulation of the VTA (VTA/tone-paired animals; see Methods). Tone pulses preceded VTA stimulation by either 925 ($n = 3$) or 500 ms ($n = 4$). The auditory cortex was mapped in detail 24 h after the last pairing session. Because no difference was observed between the auditory cortex in animals with 925 ms and 500 ms inter-stimulus pairing intervals, data from all seven animals were pooled. Compared with naïve controls paired animals had a larger physiologically defined auditory cortex (Fig. 1a, b; naïve: $1.24 \pm 0.09 \text{ mm}^2$; paired: $1.54 \pm 0.11 \text{ mm}^2$, $P < 0.05$; see Methods). In addition, a larger part of the tone-responsive auditory cortex represented the paired (9 kHz) stimulus frequency (in a range of ± 0.3 octave; $P < 0.005$). Representations of nearby frequencies (best representing 0.6 octave frequency ranges centred at 5.9 kHz and 13.6 kHz) were reduced ($P < 0.001$; Fig. 2a, b).

Representations of more spectrally distant frequencies were unaltered.

These positive and negative frequency-specific representational changes resulted in emergent sharp transitions in best frequency maps (Fig. 1b, indicated by the arrows; also see Figs 1e, f and 2a). Pairing procedures did not change response properties such as latency or the number of spikes evoked by tone presentation ($P > 0.1$ for both measures; see Methods for details). On the other hand, response bandwidths at 10 and 30 dB above threshold were reduced for neurons with best frequencies near 9 kHz ($P < 0.01$; see Fig. 2c), that is, there was an increase in spectral selectivity induced by pairing of tone pulses and VTA microstimulation. Furthermore, more neurons in experimental animals exhibited non-monotonic rate-level functions in VTA/tone-paired animals ($P < 0.005$; see Figs 1f and 2d). These non-monotonic neurons may provide a basis for intensity selectivity, and may reflect another dimension of plastic remodelling that was selective for the constant-loudness paired tonal stimuli.

To ensure that effects were attributable to the activation of dopamine neurons in the VTA (or in VTA targets), three animals received systemic SCH-23390 and eticlopride (D1- and D2-type receptor antagonists) 30 min before pairing 9-kHz pulsed tones with VTA stimulation. After 20 pairing sessions identical to those used in other experimental animals the auditory cortices of these animals were indistinguishable from those of naïve controls. Neither the sizes of auditory cortex nor the percentages of the cortical areas representing the 9-kHz stimulus frequency differed between dopamine receptor antagonist and naïve rat groups (size: $1.23 \pm 0.11 \text{ mm}^2$, $P > 0.5$ antagonist versus naïve, $P < 0.05$ antagonist versus paired, $P < 0.05$ all three groups; percentage of 9-kHz representation: $23.4 \pm 4.1\%$ for naïve group, $18.1 \pm 1.9\%$ for antagonist group, $P > 0.1$; data not shown). Dopamine neuron activity seemed to be required for the induction of cortical reorganization resulting from pairing VTA and tonal stimulation. To control for tonal exposure and VTA stimulation effects four rats were presented with 9-kHz pulsed tones and two rats received VTA stimulation for 20 sessions. No significant changes in cortical maps or receptive fields were observed in these animals (analysis of variance on total tone-responsive area, $P > 0.5$; on per cent cortex with best frequency in each frequency band, $P > 0.1$; see Supplementary Information).

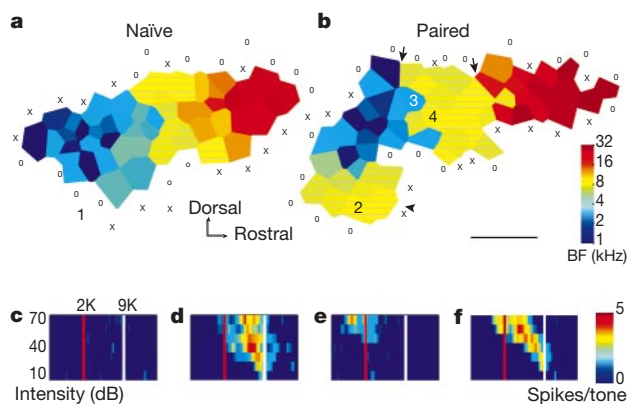


Figure 1 Representative cortical tonotopic best frequency maps and frequency-intensity receptive fields. **a, b**, Cortical maps from a naïve (**a**) and a VTA/tone-paired animal (**b**) (9-kHz pulsed tone). Hatched areas have best frequencies (BF) within 0.3 octaves of 9 kHz. Scale bar, 500 μm . O, unresponsive site; X, non-AI site. The tone-responsive auditory cortex of the paired animal consists of two separate zones: AI and a ventroposterior field (arrowhead). In AI, the representation of 9 kHz was expanded while representations of adjacent frequencies were reduced, creating sharp best-frequency transitional boundaries (arrows). **c–f**, Receptive fields recorded from sites marked 1–4 in the two representative maps.

In reorganized maps from VTA/tone-paired animals two cortical zones developed an exaggerated response to the tonal stimulus: an AI zone with best frequencies organized in a posterior–anterior direction, and a ventroposterior field responding almost exclusively to the paired stimulus frequency (Fig. 1b, indicated by an arrowhead). In paired animals, neurons in this ventroposterior field and in AI could not be easily physiologically distinguished from each other. The fact that the ventroposterior field did not conform to the posterior–anterior-oriented rising best frequency gradient—a defining feature of the rat AI (ref. 14)—indicates that it is probably not part of AI itself. This ventroposterior zone was poorly responsive to tonal stimuli and non-selective for tone frequency in naïve animals. It becomes sharply and almost exclusively tuned to the paired stimulus frequency in VTA-stimulated animals. The position of the ventroposterior zone indicates that it is Zilles' area TE2 (ref. 15), a multimodal association area^{15,16}. Dense dopamine-mediated innervation has been recorded in this region^{7,8}.

We separately documented the pairing-induced changes within these two remodelled tone-responsive auditory cortex zones, defining the borders between the presumptive AI and ventroposterior field by the lines of greatest best frequency transition. By that analysis, the size of AI alone was not significantly changed by the VTA/tone-pairing (naïve: $1.24 \pm 0.09 \text{ mm}^2$; paired: $1.21 \pm 0.08 \text{ mm}^2$, $P > 0.5$; data not shown). On the other hand, the 9-kHz representational zone in AI increased from $23.4 \pm 4.1\%$ of AI in naïve animals to $33.4 \pm 1.8\%$ in VTA/tone-paired animals ($P < 0.05$). Again, adjacent 5.9- and 13.6-kHz representations were reduced ($P < 0.05$).

The functional characteristics of a cortical neuron are defined not only by its selectivity to input parameters but also by its interactions with other neurons within cooperative neuronal ensembles. One simple measure of these interactions is the degree of synchronization of neuronal activity¹⁷. We analysed correlation of spontaneous multiunit activity in naïve ($n = 4$) and VTA/tone-paired rats ($n = 3$). In naïve animals correlation of discharges was higher for pairs of neurons in AI than for pairs in the surrounding belt region (including the ventroposterior field) or between AI and the belt

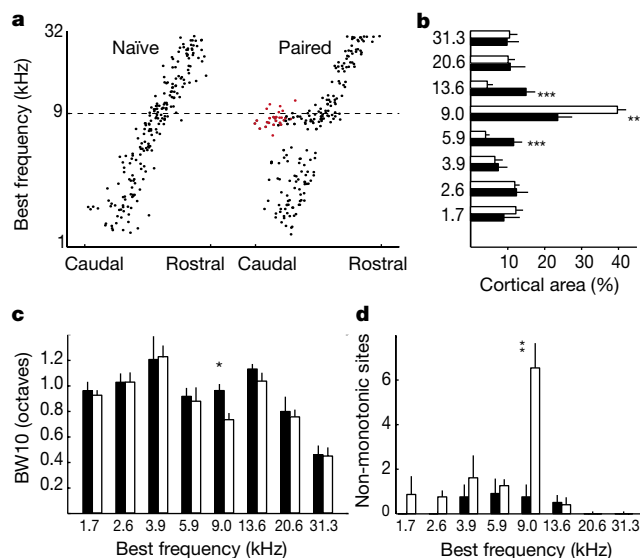


Figure 2 Reorganization of the auditory cortex by VTA microstimulation paired with a 9-kHz pulsed tone. **a**, Distribution of best frequency along the anterior–posterior axis of the auditory cortex. Points corresponding to the sites in the ventroposterior field are indicated in red ($n = 4$ for each group). **b**, Per cent of the auditory cortex that was tuned to each frequency. Black bar, naïve; white bar, paired; bin size, 0.6 octave. **c**, Response bandwidth at 10 dB (BW10) above threshold. **d**, Number of penetrations with non-monotonic rate-level function. For **b–d** $n = 6$ for the naïve group and $n = 7$ for the paired group. Asterisk, $P < 0.01$; double asterisk, $P < 0.005$; triple asterisk, $P < 0.001$.

region (collectively called non-AI pairs; data not shown). In general, correlation strengths decreased as a regular function of cortical distance for both AI and non-AI pairs. VTA/tone-pairing did not change the correlation–distance function of AI pairs, but resulted in a strong increase in the correlation strengths of neurons within the ventroposterior field and between the ventroposterior field and AI (Fig. 3a, b). In naïve animals discharge correlation in periods of spontaneous activity for VP–AI pairs was not different from that of other non-AI pairs separated by equivalent cross-cortical distances. However, in experimental animals VP–AI pairs showed much greater spontaneous activity correlation than did other non-AI pairs. This effect was specific for the cortical zones with altered plasticity—only activity from the AI area that represented the paired stimulus frequency (9-kHz AI) became highly correlated with activity from the emergent 9-kHz-representing ventroposterior zone ($P < 0.001$, between red dots and black circles in Fig. 3a, with distance ranging from 700 to 1,600 μm). Notably, these two regions were commonly more than 1 mm apart and were always separated by areas in which discharges were not measurably correlated with discharges of neurons within the ventroposterior zone (Fig. 3b). More AI pairs (83 and 80% for naïve and paired rats, respectively) showed time lags of near zero ($< 3\text{ms}$), as compared with non-AI pairs (59 and 51% for naïve and paired animals, respectively, excluding VP/9-kHz-AI pairs, Kolmogorov–Smirnov test; $P < 0.05$). The time lag distribution of the VP/9-kHz-AI pairs was comparable to that for AI pairs (82% near-zero time lag, Kolmogorov–Smirnov test; $P > 0.1$).

These results indicate that VTA dopamine neuron activity paired with sensory stimulation can induce long-range, cross-area-

synchronization of neuronal activity. They suggest that the VTA dopamine system may mediate the learning-based grouping of cortical neurons into distributed functional neuronal assemblies¹⁷.

Dopamine is believed to be involved in reinforcement learning^{4–6} in which a reward is associated with preceding but not following events⁵. We tested whether a similar temporal specificity can be obtained for this tone/VTA-pairing-induced cortical plasticity. In three rats VTA activity was preceded by a 4-kHz pulsed tone and followed by a 9-kHz pulsed tone. After 20 days of pairing, the cortical representation of 4 kHz was significantly expanded ($P < 0.0005$). The representations of adjacent 9-kHz frequencies were significantly reduced ($P < 0.05$; see Fig. 4). In addition, the ventroposterior field became well tuned to 4 kHz but not 9 kHz tones (Fig. 4). Thus, cortical reorganization enabled by VTA activity seemed to selectively enhance the saliency of the stimuli that consistently preceded and, therefore, predicted the VTA activity, and also to selectively attenuate the saliency of a stimulus that immediately followed VTA activity.

The specific mechanisms by which the dopamine system modulates reorganization of the cortex are unclear. The dense distribution of dopamine receptors in the superficial and deep layers of the cortex suggests that the dopamine system may be involved in the modulation of strengthening and weakening corticocortical connections^{7–9}. This is consistent with our findings that dopamine activity paired with sensory stimuli can enhance long-range, in-phase neuronal synchronization, which is mediated by corticocortical connections^{18,19}. Cooperative interactions between the ventroposterior field and the AI plausibly contribute to the receptive field and map changes recorded in these two zones.

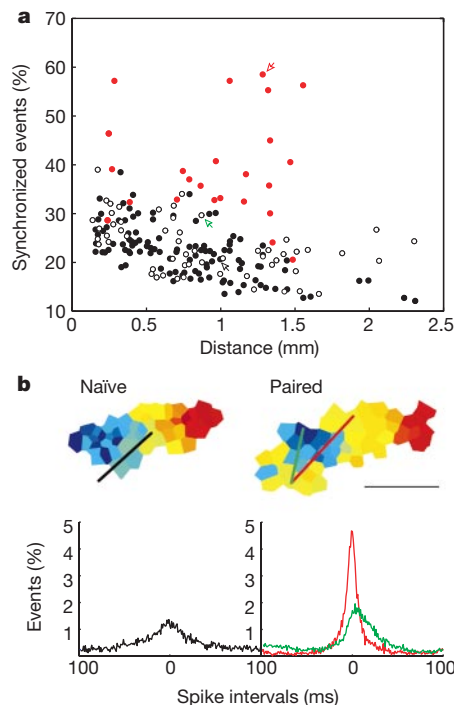


Figure 3 VTA/tone-pairing increases neuronal synchrony between the ventroposterior region and the 9-kHz-representing AI area. **a**, Synchronization–distance functions involving at least one penetration outside AI. Data from naïve animals are plotted as black dots. Data from VTA-stimulated animals were separated into two groups: pairs within the ventroposterior field and pairs between the ventroposterior field and the 9-kHz-AI area are plotted as red dots; the rest are plotted as open circles. **b**, Examples of correlograms. Black, VP/AI pair from a naïve animal; red, VP/9-kHz-AI pair from a VTA-stimulated animal; green, VP/non-9-kHz-AI pairs from the same VTA-stimulated animal. The locations of the recording sites in the best frequency maps are shown in the insets and the corresponding data points are indicated in **a** with arrows. Scale bar, 1 mm.

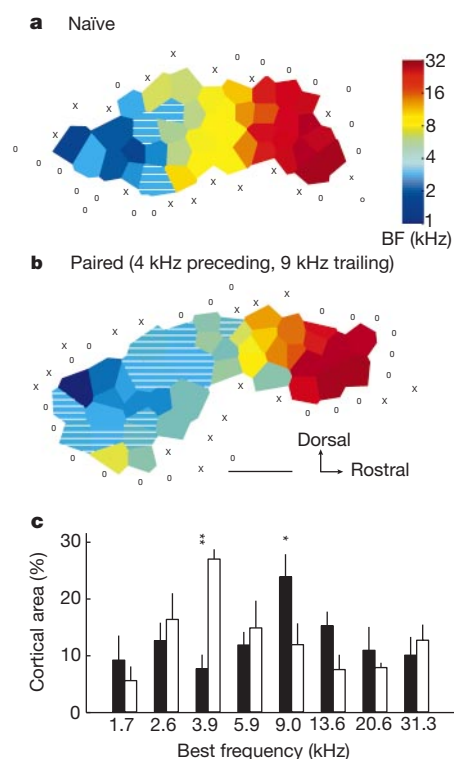


Figure 4 VTA stimulation effects are temporally asymmetrical. **a**, Representative cortical best frequency (BF) map from a naïve animal. **b**, Cortical map of an animal that received VTA stimulation paired with two pulsed tones. The first 4-kHz tone preceded the VTA stimulation by 500 ms; the second 9-kHz tone followed the VTA stimulation 500 ms later. Neurons recorded in hatched areas had best frequencies within 0.3 octave of 4 kHz. Scale bar, 500 μm . O, unresponsive site; X, non-AI site. **c**, Per cent of auditory cortex that was tuned to each frequency ($n = 6$ for naïve group and $n = 3$ for paired group). Asterisk, $P < 0.05$; double asterisk, $P < 0.0005$.

Table 1 Characteristics of auditory cortical plasticity

	VTA	Nucleus basalis*
Size of functional auditory cortex	Increased	Increased
Size of functional AI	No change	Increased
Stimulus frequency representation	Increased	Increased
Adjacent frequency representation	Decreased	Increased
Spectral selectivity	Increased	No change
Non-monotonic responses	Increased	No change
Frequency specificity of the effects	Sharper	Broader
Tuning of secondary auditory cortex	Yes	No
Temporal asymmetry of the effects	Yes	No
Modulation of stimulus-following rate	Undetermined	Yes
Cross-area synchronization	Yes	Does not apply

Auditory cortical plasticity was induced by pairing brain microstimulation with a tonal stimulus.

*Values of nucleus basalis are from refs 14, 22 and 23.

It should be emphasized that our dopamine receptor-blocking controls do not elucidate specific mechanisms accounting for cortically induced change. VTA neurons project complexly to other neuromodulatory nuclei²⁰, and indirect plasticity modulation effects may also have occurred. One of those indirect targets of the VTA is the nucleus basalis^{20,21}. Studies have already shown that powerful plastic changes can be induced in the AI in adult rats by pairing tonal stimulation with nucleus basalis stimulation following an experimental model that is similar to that applied in our study. On the other hand, changes recorded in VTA/tone-paired rats was different, in many respects, from the plasticity induced by nucleus basalis stimulation/tone pairing^{14,22,23} (see Table 1). The marked differences between VTA and nucleus basalis plasticity modulation show that changes recorded with VTA stimulation are not dominated by, and probably not substantially attributable to, indirectly engaged acetylcholine-mediated neuromodulatory inputs from the nucleus basalis. It should also be noted that VTA stimulation differs from natural activation of dopamine neurons in that it does not reflect an error in reward prediction.

The cortex is thought to be hierarchically organized, with progressively more complex information processing contributed by 'higher-level' cortical areas²⁴. Most studies have focused on plasticity in primary sensory and motor cortices and on its involvement in shaping sensory processing or motor control capabilities^{11,13,14,22,23,25–29}. In the present study, as has been indicated in earlier classical conditioning experiments³⁰, we have shown that a secondary auditory cortical zone is also subject to large-scale plasticity-induced reorganization. The 'association' cortex that has been altered by VTA/tone-pairing in the current studies has been described as diversely connected with sensory, limbic and paralimbic cortical areas, suggesting that it is involved in cross-modal sensory integration, association of emotion with sensory stimuli, and learning¹⁶. These and other studies strongly indicate that dopamine-enabled reorganization of this prospective association cortical area may have an important role in the progressive development of those cortical functions.

Our results directly demonstrate that ventral tegmental dopamine-mediated activity enables the reorganization of the cerebral cortex. The critical roles of central neuromodulatory systems in maintaining and shaping cerebral cortex network, and thereby cortical function, are just beginning to be understood. Further studies should now rapidly advance our understanding of the specific roles of different neuromodulatory systems as critical agents contributing powerfully to the progressive development of the functional capacities of the brain. □

Methods

Preparation

Platinum bipolar-stimulating electrodes were stereotactically implanted within the right VTA (4.5 mm posterior, 0.7 mm lateral and 8.5 mm ventral to Bregma) in female rats anaesthetized with barbiturate (300 g), using techniques approved under University of California, San Francisco Animal Care Facility protocols. After a two-week recovery

period rats were placed in an operant conditioning chamber and were allowed to bar press for brief VTA microstimulation (10 biphasic pulses of 0.1-ms duration at 100 Hz). The minimal current levels that reinforced consistent bar presses at least once every 2 s were determined as the electrical stimulus threshold (100–200 μ A). Subsequent pairing of auditory stimuli with VTA stimulation took place in a sound-attenuation chamber. One group of seven rats was presented with paired 9-kHz pulsed tone (six 25-ms tone pips with 5-ms on/off ramp delivered at a rate of 10 pips s^{-1} , 55 dB peak intensity) and VTA electrical microstimulation (10 biphasic pulses of 0.1-ms duration at 100 Hz, started 500 ms after tone onset). The intervals between successive pairing trials were pseudorandom in the range from 12 to 28 s. Three animals in this group were also given systemic D1 and D2 receptor antagonists (SCH-23390 and eticlopride, 0.1 and 0.3 mg kg, respectively, administered intraperitoneally 30 min before each daily pairing session). A second group of three animals underwent the same pairing procedure as the first group, except that the VTA stimulation trailed the pulsed tone by 925 ms. A third group of three animals received pairings of VTA stimulation with pulsed tones of different frequencies: a 4-kHz pulsed tone preceded the VTA stimulation and a 9-kHz pulsed tone followed it. The onset intervals between each of the tones and the onset of the VTA stimulation were both 500 ms. To control for stimulus-induced changes in the auditory cortex four animals were presented with 9-kHz pulsed tones and two animals with VTA stimulation alone.

Electrophysiology

In six naïve rats and 24 h after the last stimulation session in the experimental and control rats, animals were anaesthetized with sodium pentobarbital, the right auditory cortex surgically exposed, and neuronal responses recorded with parylene-coated tungsten microelectrodes. We chose penetration sites to evenly sample from the auditory cortical zone while avoiding blood vessels. At every penetration site the recording microelectrode was lowered orthogonal to the surface 470–550 μ m in depth (layers 4/5), where vigorous driven responses were recorded. We collected the evoked spikes of a neuron or a small cluster of 2–5 neurons at each site. Frequency/intensity response areas were reconstructed in detail by presenting 60 pure-tone frequencies (0.5–30 kHz, 25-ms duration, 5-ms ramps) at each of eight sound intensities to the contralateral ear at a rate of two stimuli per s using a calibrated sound-delivering system. The tuning curve characterization was made with a 'blind' procedure. A best frequency (recorded objectively) was defined as the frequency that evoked a neuronal response at the lowest stimulus intensity. To generate best frequency maps, points on the cortex were assigned the best frequencies of nearest penetrations through Voronoi tessellation³¹. The boundaries of the map were functionally determined using sites that did not have a well defined, tone-evoked receptive field. The response latency was defined as the time from stimulus onset to the earliest response for any of the eight intensities of the five frequencies that were nearest the best frequency. Response amplitude was defined as the average number of spikes per tone for five frequencies nearest the best frequency at 70 dB. Response latencies and amplitudes were analysed only for sites with a well defined receptive field.

To analyse neuronal discharge correlation, spontaneous activity was simultaneously recorded from three to four sites for 30 periods of non-stimulus spontaneous activity that were 2 s in duration. For each recording pair a cross-correlogram (a histogram of between-site spike intervals) was constructed from –100 ms to 100 ms. An interval less than 10 ms was considered as a synchronized event. The degree of synchronization was assessed using per cent of synchronized events. Time lag was defined as the spike interval at the peak of the correlogram. Unless otherwise specified in the text statistical significance was assessed using a two-tailed *t*-test. Data are presented as mean $\pm$ s.e.m. On completion of the experiment electrolytic lesions were made through the stimulating electrodes and the placement of the stimulating electrodes in the VTA was verified with histological examinations.

Received 28 March; accepted 2 May 2001.

- Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
- Sanes, J. N. & Donoghue, J. P. Plasticity and primary motor cortex. *Annu. Rev. Neurosci.* **23**, 393–415 (2000).
- Foote, S. L. & Morrison, J. H. Extrathalamic modulation of cortical function. *Annu. Rev. Neurosci.* **10**, 67–95 (1987).
- Goldsmith, S. K. & Joyce, J. N. Dopamine D2 receptors are organized in bands in normal human temporal cortex. *Neurosci.* **74**, 435–451 (1996).
- Lidow, M. S., Goldman-Rakic, P. S., Gallager, D. W. & Rakic, P. Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using (³H)raclopride, (³H)spiperone and (³H)SCH23390. *Neuroscience* **40**, 657–671 (1991).
- Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* **1**, 304–309 (1998).
- Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
- Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
- Stark, H. & Scheich, H. Dopaminergic and serotonergic neurotransmission systems are differentially involved in auditory cortex learning: a long-term microdialysis study of metabolites. *J. Neurochem.* **68**, 691–697 (1997).
- Bakin, J. S., South, D. A. & Weinberger, N. M. Induction of receptive field plasticity in the auditory cortex of the guinea pig during instrumental avoidance conditioning. *Behav. Neurosci.* **110**, 905–913 (1996).
- Gurden, H., Takita, M. & Jay, T. M. Essential role of D1 but not D2 receptors in the NMDA receptor-dependent long-term potentiation at hippocampal-prefrontal cortex synapses *in vivo*. *J. Neurosci.* **20**, RC106, 1–5 (2000).
- Otmakhova, N. A. & Lisman, J. E. D1/D5 dopamine receptors inhibit depotentiation at CA1 synapses via cAMP-dependent mechanism. *J. Neurosci.* **18**, 1270–1279 (1998).

13. Rioult-Pedotti, M.-S., Friedman, D. & Donoghue, J. P. Learning-induced LTP in neocortex. *Science* **290**, 533–536 (2000).
14. Kilgard, M. P. & Merzenich, M. M. Cortical map reorganization enabled by nucleus basalis activity. *Science* **279**, 1714–1718 (1998).
15. Zilles, K. in *The Cerebral Cortex of the Rat* (eds Kolb, B. & Tees, R. C.) 77–112 (MIT, Cambridge, Massachusetts, 1990).
16. Pandya, D. N. & Seltzer, B. Association areas of the cerebral cortex. *Trends Neurosci.* **5**, 386–390 (1982).
17. Singer, W. Neuronal synchrony: a versatile code for the definition of relations? *Neuron* **24**, 49–65 (1999).
18. Engel, A. K., König, P., Kreiter, A. K. & Singer, W. Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* **252**, 1177–1179 (1991).
19. Traub, R. D., Whittington, M. A., Stanford, I. M. & Jefferys, J. G. A mechanism for generation of long-range synchronous fast oscillations in the cortex. *Nature* **383**, 621–624 (1996).
20. Oades, R. D. & Halliday, G. M. Ventral tegmental (A10) system: neurobiology. 1. Anatomy and connectivity. *Brain Res.* **434**, 117–165 (1987).
21. Haring, J. H. & Wang, R. Y. The identification of some sources of afferent input to the rat nucleus basalis magnocellularis by retrograde transport of horseradish peroxidase. *Brain Res.* **366**, 152–158 (1986).
22. Kilgard, M. P. & Merzenich, M. M. Plasticity of temporal information processing in the primary auditory cortex. *Nature Neurosci.* **1**, 727–731 (1998).
23. Bakin, J. S. & Weinberger, N. M. Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proc. Natl Acad. Sci. USA* **93**, 11219–11224 (1996).
24. Van Essen, D. C., Anderson, C. H. & Felleman, D. J. Information processing in the primate visual system: an integrated systems perspective. *Science* **255**, 419–423 (1992).
25. Tremblay, N., Warren, R. A. & Dykes, R. W. Electrophysiological studies of acetylcholine and the role of the basal forebrain in the somatosensory cortex of the cat. II. Cortical neurons excited by somatic stimuli. *J. Neurophysiol.* **64**, 1212–1222 (1990).
26. Sachdev, R. N., Lu, S. M., Wiley, R. G. & Ebner, F. F. Role of the basal forebrain cholinergic projection in somatosensory cortical plasticity. *J. Neurophysiol.* **79**, 3216–3228 (1998).
27. Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, T. & Guic-Robles, E. Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *J. Neurophysiol.* **63**, 82–104 (1990).
28. Recanzone, G. H., Schreiner, C. E. & Merzenich, M. M. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* **13**, 87–103 (1993).
29. Nudo, R. J., Wise, B. M., SiFuentes, F. & Milliken, G. W. Neural substrates for the effects of rehabilitative training on motor recovery after ischemic infarct. *Science* **272**, 1791–1794 (1996).
30. Diamond, D. M. & Weinberger, N. M. Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. *Brain Res.* **372**, 357–360 (1986).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Blake for many informative suggestions; L. I. Zhang and F. Strata for discussions; and C. Garabedian and T. Moallem for comments on the manuscript. This work was supported by an NIH grant, the Coleman Fund, the Mental Insight Foundation, and Hearing Research.

Correspondence and requests for materials should be addressed to M.M.M. (e-mail: merz@phy.ucsf.edu).

The glyoxylate cycle is required for fungal virulence

Michael C. Lorenz & Gerald R. Fink

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

***Candida albicans*, a normal component of the mammalian gastrointestinal flora, is responsible for most fungal infections in immunosuppressed patients. *Candida* is normally phagocytosed by macrophages and neutrophils, which secrete cytokines and induce hyphal development in this fungus^{1,2}. Neutropenic patients, deficient in these immune cells, are particularly susceptible to systemic candidiasis^{3,4}. Here we use genome-wide expression profiles of the related yeast *Saccharomyces cerevisiae* to obtain a signature of the events that take place in the fungus on ingestion by a mammalian macrophage. Live *S. cerevisiae* cells isolated from**

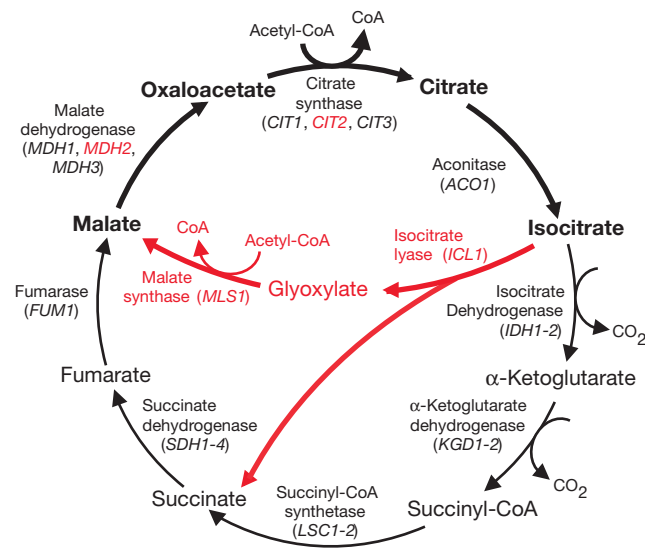
the phagolysosome are induced for genes of the glyoxylate cycle, a metabolic pathway that permits the use of two-carbon compounds as carbon sources. In *C. albicans*, phagocytosis also up-regulates the principal enzymes of the glyoxylate cycle, isocitrate lyase (ICL1) and malate synthase (MLS1). *Candida albicans* mutants lacking *ICL1* are markedly less virulent in mice than the wild type. These findings in fungi, in conjunction with reports that isocitrate lyase is both upregulated and required for the virulence of *Mycobacterium tuberculosis*^{5,6}, demonstrate the wide-ranging significance of the glyoxylate cycle in microbial pathogenesis.

Systematic studies of host–pathogen interactions have been hampered by the lack of genetic tools in *C. albicans*. For this reason the related but non-pathogenic yeast *S. cerevisiae* is often used to uncover relevant genes. *In vitro*, cultured mammalian macrophages readily ingest both *S. cerevisiae* and *C. albicans* cells. A population of *S. cerevisiae* highly enriched for phagocytosed cells was isolated and subjected to whole-genome microarray analysis using oligonucleotide-based arrays (Affymetrix). Three hours after initiating the co-culture, most of the phagocytosed cells were alive (averaging 67% alive, as assayed by methylene blue staining); transcriptional profiling of these cells reveals the response of fungal cells to phagocytosis.

Eleven of the fifteen most highly induced *S. cerevisiae* genes after phagocytosis (Table 1) encode proteins related to the glyoxylate cycle, through which two-carbon compounds are assimilated into the tricarboxylic acid (TCA) cycle (see Fig. 1). Three of the five glyoxylate cycle enzymes are on this list (isocitrate lyase, ICL1; malate synthase, MLS1; and malate dehydrogenase, MDH2), and a fourth (citrate synthase, CIT2) is also strongly induced (4.9-fold, ranking 24th). Furthermore, several genes functionally related to the glyoxylate cycle are induced, including acetyl coenzyme A (acetyl-CoA) synthase (ACS1); YDR384c, a homologue of the *Yarrowia lipolytica* glyoxylate pathway regulator (GPR1; refs 7, 8); several transporters and acetyltransferases, which are used to traffic intermediates of the glyoxylate cycle and fatty-acid degradation between organelles (CRC1, ACR1, YAT1 and YER024w); and fructose-1,6-bisphosphatase (FBP1). FBP1 is a central regulatory point in gluconeogenesis⁹—the production of glucose is the principal function of the glyoxylate cycle. Induction of the glyoxylate cycle indicates that nutrient acquisition and use is the primary focus of yeast cells upon phagocytosis, presumably because the phagolysosome is poor in complex carbon compounds (see Fig. 1).

Although the glyoxylate cycle and TCA share common reactions, it is only the isozymes specialized for the glyoxylate cycle that are induced (Fig. 1). The cytosolic isozyme of MDH2, which preferentially functions in the glyoxylate cycle¹⁰, is induced 15.6-fold. By contrast, the mitochondrial (MDH1) and peroxisomal (MDH3) forms are not induced. Out of the three citrate synthase isoforms only the glyoxylate cycle-specific CIT2 is induced. In control array experiments expression of glyoxylate cycle enzymes were not changed significantly in response to conditioned media, oxidative stress, or contact with heat-killed macrophages (see Methods). Thus, phagocytosis specifically upregulates the glyoxylate cycle and its accessory proteins. This metabolic response takes precedence over any conventional stress response, suggesting that nutrient deprivation is the primary 'stress' that confronts these cells.

We cloned the *C. albicans* genes for isocitrate lyase (*ICL1*) and malate synthase (*MLS1*), the only enzymes whose activity is both specific and limited to the glyoxylate cycle. Both genes share significant homology with proteins from fungi, plants and bacteria but, notably, not mammals, which do not have the glyoxylate cycle. Northern analysis of RNA from both *S. cerevisiae* and *C. albicans* cells grown in the presence of macrophages shows that in both organisms the *ICL1* or *MLS1* (Fig. 2a, b) genes are significantly induced by macrophage contact when compared with cells grown in media alone. Thus, the induction of the glyoxylate enzymes is a



Gene	Induction	Gene	Induction
<i>MLS1</i>	22.7	<i>SDH3</i>	1.0
<i>ICL1</i>	22.4	<i>SDH2</i>	0.9
<i>MDH2</i>	15.6	<i>ACO1</i>	0.8
<i>CIT2</i>	4.9	<i>MDH3</i>	0.8
<i>SDH1</i>	1.9	<i>MDH1</i>	0.7
<i>SDH4</i>	1.7	<i>FUM1</i>	0.6
<i>KGD1</i>	1.4	<i>IDH1</i>	0.4
<i>CIT1</i>	1.1		

Figure 1 Schematic drawing of the glyoxylate and tricarboxylic acid cycles. The names of the corresponding *S. cerevisiae* genes are included for each enzyme. Glyoxylate-specific steps and enzymes are shown in red. The fold-induction values (for the phagocytosed population compared with those grown in media alone) for each gene are shown below. *CIT3*, *LSC1*, *LSC2*, *KGD2* and *IDH2* did not meet our filter requirements (see Methods) and are excluded from this list.

conserved response to phagocytosis in these two yeasts, which diverged from a common ancestor an estimated 150 million years ago.

We constructed mutant strains of both *S. cerevisiae* and *C. albicans* that lacked *ICL1*. In both organisms the *icl1* mutant strains fail to use acetate or ethanol, as expected (Fig. 3a; and data not shown). In *C. albicans*, both the heterozygous strain ($\Delta icl1/ICL1$) and a homozygous mutant in which *ICL1* has been re-introduced ($\Delta icl1/\Delta icl1 + ICL1$) grow as well as a wild-type strain on acetate media (Fig. 3a, b). The growth rate of the *C. albicans* $\Delta icl1/\Delta icl1$

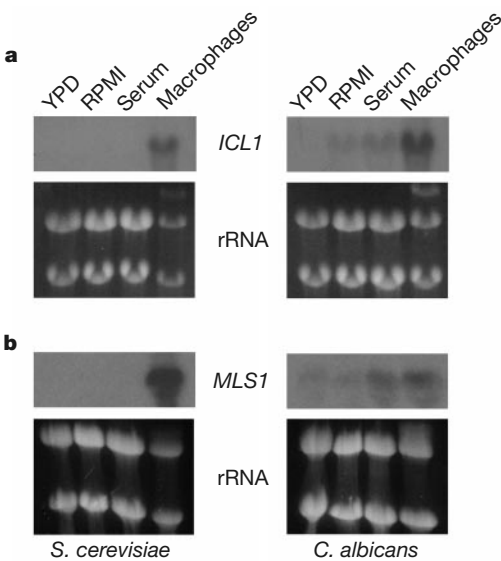


Figure 2 Macrophage contact induces isocitrate lyase and malate synthase in both *S. cerevisiae* and *C. albicans*. **a, b**, RNA prepared from yeast cells exposed to macrophages for 3 h and hybridized to radiolabelled probes made from *S. cerevisiae* or *C. albicans* *ICL1* (**a**) or *MLS1* (**b**).

strain is not significantly different from the parent strain on rich (YP-Dextrose) media (Fig. 3b), nor is this strain any more sensitive to a variety of *in vivo* stresses, including salt, heat shock, ethanol (assayed on glucose media), or oxidative stress (Fig. 3c, d). $\Delta icl1/\Delta icl1$ strains form filaments (on solid medium) and germ tubes (in liquid medium) in response to serum or neutral pH (Fig. 3e; and data not shown).

We tested the virulence of these *C. albicans* strains in a mouse model of systemic candidiasis. Mice injected with wild-type *C. albicans* strain SC5314 succumb rapidly to the infection (median survival of 3 days; Fig. 4), whereas mice injected with two independently constructed $\Delta icl1/\Delta icl1$ strains survived longer. At day 28, 7 out of 10 of the animals injected with one strain (MLC7) remained alive as did 6 out of 10 of an independent homozygous mutant (MLC8). Infection with the heterozygote ($\Delta icl1/ICL1$) resulted in an intermediate mortality (median of 8 days). Thus, isocitrate lyase is not only induced by macrophage phagocytosis but is also essential for full virulence in this fungal pathogen.

Our data from genome arrays and virulence studies suggest that microbes find the inside of a macrophage to be a glucose-deficient environment. Glucose is required for the synthesis of many macromolecules necessary for proliferation, including ribose and

Table 1 Genes induced by phagocytosis in *S. cerevisiae*

Gene	ORF	YPD	Serum	Macrophage	Description
<i>YAT1</i>	YAR035W	0.9	1.0	36.6	Outer carnitine acetyl transferase; mitochondrial
ORF	YMR031C	5.1	1.0	36.1	Unknown
<i>ICL1</i>	YER065C	0.2	1.0	22.7	Isocitrate lyase, peroxisomal (glyoxylate cycle)
<i>MLS1</i>	YNL117W	0.2	1.0	22.5	Malate synthase (glyoxylate cycle)
<i>MDH2</i>	YOL126C	0.3	1.0	15.7	Malate dehydrogenase, cytosolic (glyoxylate cycle)
<i>NCE3</i>	YNL036W	1.5	1.0	13.8	Similar to carbonic anhydrase; substrate for non-classical export pathway
ORF	YDR384C	0.4	1.0	12.1	Similar to <i>Y. lipolytica</i> Gpr1
ORF	YKL187C	2.2	1.0	11.6	Similar to 4-mycarosyl isovaleryl-CoA; induced by glycerol, oleate
<i>FBP1</i>	YLR377C	0.1	1.0	10.8	Fructose-1,6-bisphosphatase
ORF	YMR118C	0.2	1.0	10.3	Succinate dehydrogenase; similar to Sdh3
<i>SPS100</i>	YHR139C	0.8	1.0	9.8	Sporulation specific protein; induced by ethanol
<i>ACS1</i>	YAL054C	0.2	1.0	8.7	Acetyl-CoA synthetase
<i>CRC1</i>	YOR100C	0.1	1.0	8.1	MCF; involved in carnitine transport
<i>ACR1</i>	YJR095W	0.0	1.0	6.8	Mitochondrial succinate-fumarate transporter (MCF family)

Values are fold-induction compared to expression in tissue culture medium plus serum (Serum) for cells grown in rich medium (YPD) or the phagocytosed cell population (Macrophage). Values represent the average of duplicate array experiments. The 'Description' column is derived from the Yeast Protein Database maintained by Proteome (www.proteome.com/YPDhome.html). MCF, mitochondrial carrier family; ORF, open reading frame.

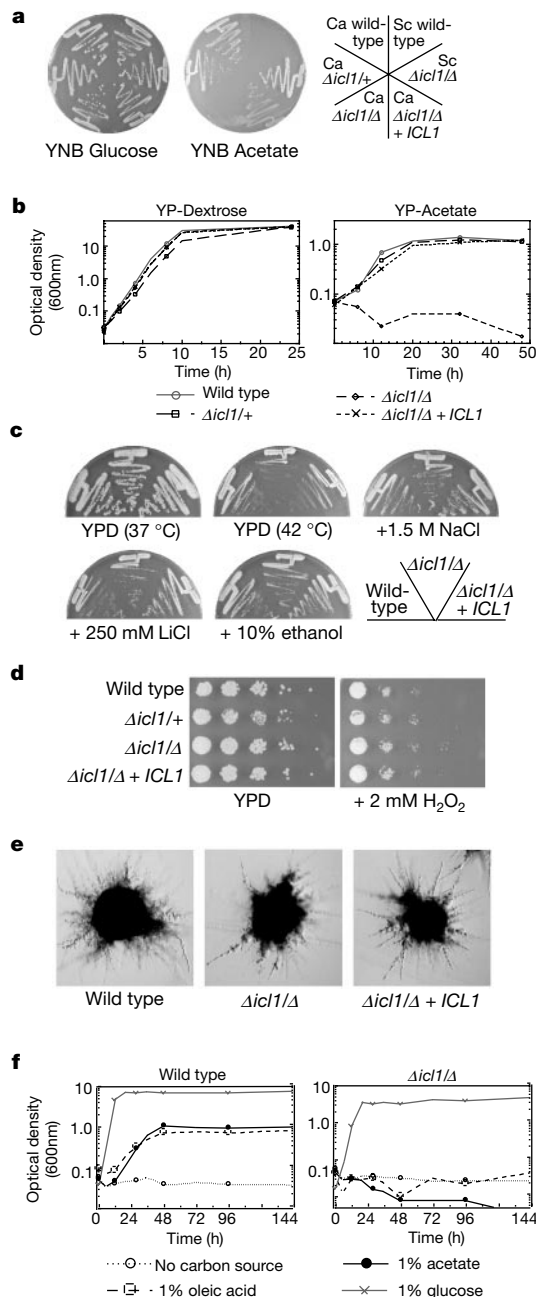


Figure 3 Phenotypes of isocitrate lyase mutant strains. $\Delta icl1/+ = \Delta icl1/ICL1$; $\Delta icl1/\Delta = \Delta icl1/icl1$. **a**, *Candida albicans* strains SC5314 (wild type), MLC6 ($\Delta icl1/ICL1$), MLC7 ($\Delta icl1/\Delta icl1$) and MLC10 ($\Delta icl1/\Delta icl1 + ICL1$), or *S. cerevisiae* strains EM93 (wild type) and MLY283 ($\Delta icl1/\Delta icl1$ MATa α) incubated on YNB media containing 2% glucose (left) or 2% sodium acetate (right) as the sole carbon source. The cells grown on glucose were incubated at 37 °C for 2 d; cells grown on acetate were incubated for 4 d. **b**, *Candida albicans* strains SC5314, MLC6, MLC7 and MLC10 grown in liquid YP-Dextrose (left) or YNB containing 2% sodium acetate as the sole carbon source (right) at 37 °C for the indicated time period. **c**, SC5314, MLC7 and MLC10 grown on YPD containing 1.5 M NaCl, 250 mM LiCl, 10% ethanol (at 37 °C), or on YPD at 42 °C for 30 h. **d**, Liquid cultures of SC5314, MLC6, MLC7 and MLC10 grown in YPD at 37 °C to an optical density at 600 nm of about 1.0, serially diluted (1:10 dilutions), and plated by spotting to media with YPD plus 5 mM H₂O₂ at 37 °C for 30 h. **e**, SC5314, MLC7 and MLC10 incubated on 2% agar/10% serum media for 30 h at 37 °C and photographed at $\times 100$ magnification. **f**, *Candida albicans* strains SC5314, MLC6, MLC7 and MLC10 grown in liquid YP-Dextrose (left) or YNB containing 2% sodium acetate as the sole carbon source (right), at 37 °C for the indicated time period. The $\Delta icl1/\Delta icl1$ strain grew to about half the final density of the wild-type strain in glucose media, perhaps indicative of a failure to grow past the diauxic shift at which yeast cells must begin to use ethanol produced during the earlier fermentative growth phase.

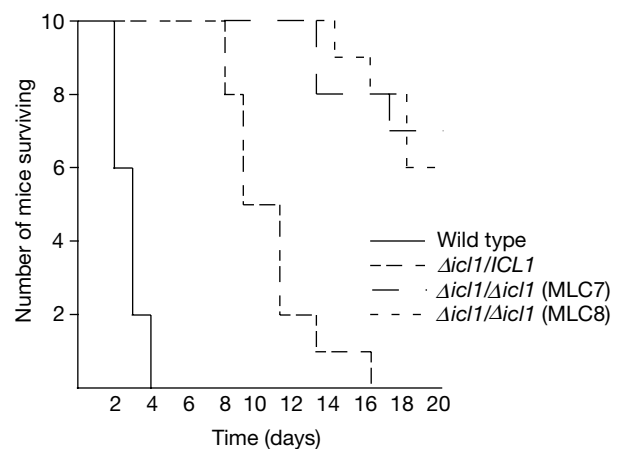


Figure 4 Isocitrate lyase mutants are avirulent. Cells (6×10^5) of the indicated strains were injected into the tail vein of BALB/c mice ($n = 10$) and monitored for 28 d. In an independent experiment, re-introducing a copy of the *ICL1* gene (MLC10; $\Delta icl1/\Delta icl1 + ICL1$) restored virulence to levels similar to the wild type.

deoxyribose. We speculate that the phagolysosome is rich in fatty acids or their breakdown products (primarily acetyl-CoA). Acetyl-CoA can only be assimilated through the glyoxylate cycle, which bypasses the catabolic steps of the TCA cycle; thus the glyoxylate cycle is the only route to the synthesis of glucose in this environment. We assayed growth of *C. albicans* in media containing oleic acid, an unsaturated 18-carbon fatty acid, as the sole carbon source (Fig. 3f). The wild-type strain uses oleic acid as well as acetate, whereas the $\Delta icl1/\Delta icl1$ strain is unable to metabolize oleic acid. Although the glyoxylate pathway is necessary for virulence it is clearly not sufficient. Both *Saccharomyces* and *Candida* induce the glyoxylate cycle on macrophage contact, yet only *Candida* is virulent.

Genes encoding the glyoxylate cycle have now been shown to be required for virulence in both a bacterium (*M. tuberculosis*) and a fungus (*C. albicans*) that can survive inside a macrophage. Inhibitors of the glyoxylate cycle pathway should block nutrient availability and prevent survival of these pathogens inside the macrophage. Compounds that inhibit nutrient availability have been developed into effective herbicides (for example, glyphosate and imidizolinones) because their targets are enzymes produced by plants but not by animals. As the enzymes of the glyoxylate cycle are also not found in mammals, they are prime targets for antibacterial and antifungal agents. □

Methods

Yeast-macrophage co-culture and gene expression analysis

Murine macrophage-like cell line J774A (ATCC number TIB-67) was cultured in RPMI plus 10% fetal bovine serum at 37 °C in 95% air/5% CO₂. About 18 h before a co-culture experiment, cells were plated in 50 ml media at 2×10^7 cells per 750 ml flask. Yeast strain EM93 (MATa α prototroph¹¹) was grown overnight in YPD media at 37 °C, then diluted in fresh YPD for 3–4 h (approximately two doublings). Yeast cells were pelleted by centrifugation, washed once, resuspended in PBS, and added to the J774A cultures at 4×10^8 cells per flask (a tenfold excess assuming that the J774A cells doubled once after plating). The co-culture was incubated for 2.5–3.0 h at 37 °C in normal air. Yeast cells not associated with the adherent macrophages were removed by washing three times with ice-cold PBS. The macrophages and associated yeast were removed by scraping, and pooled by centrifugation for 1 min at 500g. We determined cell number and viability (using methylene blue) by microscopy. The cell mixture was washed twice with ice-cold water to lyse the mammalian cells. The resulting cell pellet consisting mostly of yeast cells was frozen at –80 °C.

RNA was made from the pooled cell pellets using hot acidic phenol and the poly(A) fraction was isolated using the Poly(A)tract kit (Promega). Poly(A) RNA (2 μ g per sample) was labelled in duplicate as described¹² and hybridized to the Ye6100 oligonucleotide array set (Affymetrix). Array data were extracted and filtered to remove any genes whose expression did not change at least twofold (or 100 units) in the experiment. The complete dataset is available at http://staffa.wi.mit.edu/fink_public/glyoxylate.

Candida albicans homologues of *ICL1* and *MLS1* were identified by searching currently available *C. albicans* genome sequence data from the Stanford DNA Sequencing and

Technology Center (<http://sequence-www.stanford.edu/group/candida/index.html>). There are single homologues for each gene (the gene sequences for *C. albicans* *ICL1* and *MLS1* are deposited in GenBank under accession numbers AF222905 and AF222907, respectively). For northern analysis, macrophage interactions were performed as described above using 10^6 J774A cells in 5 ml media with 2×10^7 *S. cerevisiae* (EM93) or *C. albicans* (SC5314) cells. Control populations were grown for 3 h in rich media (YPD), or in tissue culture media without (RPMI) or with (serum) 10% fetal bovine serum. Species-specific probes were amplified by PCR and labelled with a random primer.

Mutant construction and analysis

Saccharomyces cerevisiae $\Delta icl1$ mutants were constructed in the EM93 background using a PCR-mediated protocol with a G418-resistance cassette¹³. Mutants were constructed in both mating types and mated to produce a homozygous $\Delta icl1/\Delta icl1$ knockout strain (MLY283a/α). For *C. albicans*, an $\Delta icl1$ disruption construct was created by inserting a *hisG*–*URA3*–*hisG* cassette¹⁴ at a *Bgl*II site in the *ICL1* open reading frame. This construct was linearized, transformed into CA14 (a *Ura*[−] derivative of strain SC5314; refs 14, 15), and selected by uracil prototrophy. Accurate integrants were identified by PCR and passed on 5-FOA medium. A second round of transformation was used to generate two independent homozygous $\Delta icl1/\Delta icl1$ strains (MLC7 and MLC8; MLC7 was used for most of the experiments reported here). The wild-type *ICL1* gene was re-introduced on linearized plasmid pRC2312 (ref. 16) by transformation to produce a complemented strain (MLC10). *Candida albicans* transformations were done as described¹⁷. Standard media was used¹⁸ and strains were grown at 37 °C unless otherwise indicated.

Murine virulence assay

Overnight cultures of *C. albicans* strains were diluted into fresh YPD and grown for 3–4 h at 37 °C. Cultures were collected by centrifugation and washed with PBS. Cells (6×10^5) were injected into the tail vein of 18–20-week-old female BALB/c mice. We used ten mice per strain. Mice were monitored for three weeks after injection and moribund animals were killed. We cared for animals according to NIH guidelines.

Received 9 February; accepted 8 May 2001.

- Ashman, R. B. & Papadimitriou, J. M. Production and function of cytokines in natural and acquired immunity to *Candida albicans* infection. *Microbiol. Rev.* **59**, 646–672 (1995).
- Lo, H. J. et al. Nonfilamentous *C. albicans* mutants are avirulent. *Cell* **90**, 939–949 (1997).
- Wright, W. L. & Wenzel, R. P. Nosocomial *Candida*. Epidemiology, transmission, and prevention. *Infect. Dis. Clin. North. Am.* **11**, 411–425 (1997).
- Bodey, G. et al. Fungal infections in cancer patients: an international autopsy survey. *Eur. J. Clin. Microbiol. Infect. Dis.* **11**, 99–109 (1992).
- McKinney, J. D. et al. Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735–738 (2000).
- Honer Zu Bentrop, K., Miczak, A., Swenson, D. L. & Russell, D. G. Characterization of activity and expression of isocitrate lyase in *Mycobacterium avium* and *Mycobacterium tuberculosis*. *J. Bacteriol.* **181**, 7161–7167 (1999).
- Kujau, M., Weber, H. & Barth, G. Characterization of mutants of the yeast *Yarrowia lipolytica* defective in acetyl-coenzyme A synthetase. *Yeast* **8**, 193–203 (1992).
- Tzschoppe, K., Augstein, A., Bauer, R., Kohlwein, S. D. & Barth, G. trans-dominant mutations in the *GPR1* gene cause high sensitivity to acetic acid and ethanol in the yeast *Yarrowia lipolytica*. *Yeast* **15**, 1645–1656 (1999).
- Sedivy, J. M. & Fraenkel, D. G. Fructose biphosphatase of *Saccharomyces cerevisiae*. Cloning, disruption and regulation of the *FBP1* structural gene. *J. Mol. Biol.* **186**, 307–319 (1985).
- Minard, K. I. & McAlister-Henn, L. Isolation, nucleotide sequence analysis, and disruption of the *MDH2* gene from *Saccharomyces cerevisiae*: evidence for three isozymes of yeast malate dehydrogenase. *Mol. Cell. Biol.* **11**, 370–380 (1991).
- Mortimer, R. K. & Johnston, J. R. Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**, 35–43 (1986).
- Wodicka, L., Dong, H., Mittmann, M., Ho, M. H. & Lockhart, D. J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol.* **15**, 1359–1367 (1997).
- Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**, 1793–1808 (1994).
- Fonzi, W. A. & Irwin, M. Y. Isogenic strain construction and gene mapping in *Candida albicans*. *Genetics* **134**, 717–728 (1993).
- Gillum, A. M., Tsay, E. Y. & Kirsch, D. R. Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol. Gen. Genet.* **198**, 179–182 (1984).
- Cannon, R. D., Jenkinson, H. F. & Shepherd, M. G. Cloning and expression of *Candida albicans* *ADE2* and proteinase genes on a replicative plasmid in *C. albicans* and in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **235**, 453–457 (1992).
- Braun, B. R. & Johnson, A. D. Control of filament formation in *Candida albicans* by the transcriptional repressor TUP1. *Science* **277**, 105–109 (1997).
- Sherman, F. Getting started with yeast. *Methods Enzymol.* **194**, 3–21 (1991).

Acknowledgements

We are grateful to G. Acton and E. Lander for access to microarray facilities; T. Galitski, A. Saldhana, C. Huff and F. Lewitter for help with informatics resources; and members of the Fink laboratory for useful discussions. This work was supported by NIH grants. G.R.F. is an American Cancer Society Professor of Genetics. M.C.L. is a post-doctoral fellow of the Irvington Institute for Immunological Research.

Correspondence and requests for materials should be addressed to G.R.F. (e-mail: fink@wi.mit.edu).

Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains

Charles E. Dann*, Jen-Chih Hsieh†‡, Amir Rattner†‡, Divya Sharma*, Jeremy Nathans†‡§ & Daniel J. Leahy*‡

* Department of Biophysics and Biophysical Chemistry, † Department of Molecular Biology and Genetics, ‡ Howard Hughes Medical Institute, § Departments of Neuroscience and Ophthalmology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Members of the Frizzled family of seven-pass transmembrane proteins serve as receptors for Wnt signalling proteins¹. Wnt proteins have important roles in the differentiation and patterning of diverse tissues during animal development^{2,3}, and inappropriate activation of Wnt signalling pathways is a key feature of many cancers⁴. An extracellular cysteine-rich domain (CRD) at the amino terminus of Frizzled proteins binds Wnt proteins¹, as do homologous domains in soluble proteins—termed secreted Frizzled-related proteins⁵—that function as antagonists of Wnt signalling^{6–8}. Recently, an LDL-receptor-related protein has been shown to function as a co-receptor for Wnt proteins and to bind to a Frizzled CRD in a Wnt-dependent manner^{9–11}. To investigate the molecular nature of the Wnt signalling complex, we determined the crystal structures of the CRDs from mouse Frizzled 8 and secreted Frizzled-related protein 3. Here we show a previously unknown protein fold, and the design and interpretation of CRD mutations that identify a Wnt-binding site. CRDs exhibit a conserved dimer interface that may be a feature of Wnt signalling. This work provides a framework for studies of homologous CRDs in proteins including muscle-specific kinase and Smoothened, a component of the Hedgehog signalling pathway^{12,13}.

The CRDs of secreted Frizzled-related protein 3 (sFRP-3) and mouse Frizzled 8 (mFz8), which possess 10 conserved cysteines within a domain of 120–125 amino acids, were expressed in Chinese hamster ovary (CHO) cells as a fusion with human growth hormone¹⁴. Putative N-linked glycosylation sites were eliminated by mutation, a step that proved essential to obtain diffraction-quality crystals. After cleavage of the fusion protein, mutant CRDs (CRDM) exhibited the same affinity for *Xenopus* Wnt-8 (XWnt8) as native CRDs and produced crystals that diffracted beyond 2.0 Å Bragg spacings. The structure of sFRP-3 CRDM was solved by multi-wavelength anomalous diffraction¹⁵ using the anomalous signal from bound solvent bromide ions¹⁶. Nine bromide ions provided sufficient phasing for structure determination without non-crystallographic symmetry averaging despite the presence of six molecules (relative molecular mass 84,000 (M_r 84K)) in the asymmetric unit. The structure of mFz8 CRDM was determined by molecular replacement using a truncated sFRP-3 CRDM as a search model.

The CRDM structures (Fig. 1a, b) are predominantly α -helical with all cysteines forming disulphide bonds (Cys3–Cys64, Cys11–Cys57, Cys48–Cys87, Cys76–Cys115, Cys80–Cys104 in sFRP-3). The program DALI¹⁷ and inspection of SCOP¹⁸ and CATH¹⁹ failed to identify a clear structural homologue, although visual inspection hints at a distant relationship to four-helix bundles. In addition to helical regions, two short β -strands at the N terminus form a minimal β -sheet with β_2 passing through a knot created by disulphide bonds. The sFRP-3 and mFz8 CRDs are arranged as homologous dimers within the crystallographic asymmetric units with $\sim 1,500$ Å² and ~ 900 Å² buried, respectively, at a highly complementary dimer interface (Fig. 1c, d). Shape complementarity analysis of sFRP-3 and mFz8 dimers with the program SC showed values (0.71 and 0.64, respectively) consistent with known

protein–protein interfaces²⁰. Both purified CRDs are, however, monomeric in solution at 100 μ M as judged by gel filtration chromatography, indicating a dissociation constant for the dimers between this value and 50 mM, the concentration of the CRDs in the crystals.

Several studies indicate a direct interaction between Wnt proteins and Frizzled or sFRP CRDs^{1,5,7,21}. To identify regions of the CRD important for Wnt binding, we used a binding assay²¹ *in vitro* and three mutagenesis strategies: tripeptide insertion, alanine scanning and homologue scanning²². In the binding experiments, the CRD was displayed on the surface of transiently transfected cells by a Myc-tagged glycosylphosphatidylinositol (GPI) anchor; and the cell-surface binding of a soluble *Xenopus* Wnt8–alkaline phosphatase fusion protein (XWnt8–AP) measured histochemically (Fig. 2). All of the mutants studied localized efficiently to the plasma membrane, as determined by immunostaining of living cells with anti-Myc antibody and release of the CRD–Myc fusion protein from the surface of living cells with phosphoinositide-phospholipase C. CRDs of *Drosophila* Frizzled 2 (DFz2) and mFz8 were used in these experiments because they exhibit a high affinity for XWnt8.

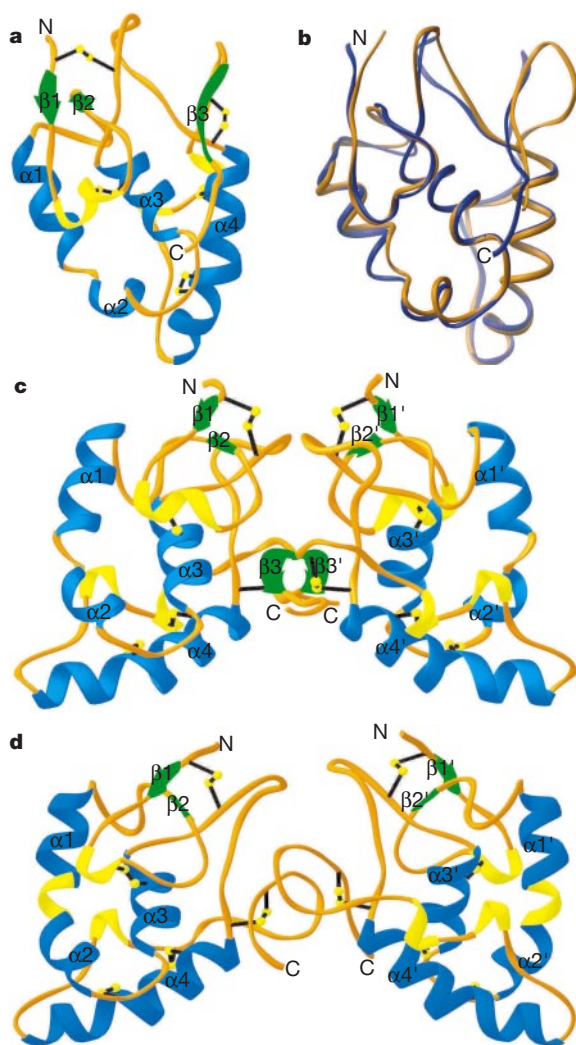


Figure 1 Crystal structure of the sFRP-3 and mFz8 CRDs. **a**, Ribbon diagram of the sFRP-3 CRD with elements of secondary structure numbered in order of appearance in the primary structure. **b**, Superposition of sFRP-3 CRD (blue) and mFz8 CRD (brown). **c**, Dimer of the sFRP-3 CRD observed in the crystal. **d**, Dimer of the mFz8 CRD observed in the crystal. **a**, **c**, **d**, α -helices, blue; β -strands, green; coil, gold. Disulphide bonds are shown as ball-and-stick models in yellow and black. Images were generated by RIBBONS²⁸.

Given their homologous structure and function, we presume that all Frizzled CRDs will bind Wnt proteins in a conserved fashion.

In earlier efforts to analyse Wnt–CRD binding, a series of 23 tripeptide insertions were constructed in the DFz2 CRD²¹. Guided by the sFRP-3 crystal structure, 11 additional tripeptide insertions were constructed in the DFz2 CRD to complete coverage of the CRD surface. The insertion sites for the entire set of 34 mutants and a summary of their effects on Wnt binding are shown in Fig. 3. The 16 insertion mutants that retain XWnt8–AP binding most likely reside in regions that do not make direct Wnt–CRD contacts (Fig. 3, green arrowheads; Fig. 4a, green areas). In general, members of this class of insertions were located within surface loops. Interpretation of the 18 insertion mutants that disrupt binding (Fig. 3, red arrowheads) is less clear. Whereas some insertions may directly block binding, many reside in regions of secondary structure and could affect binding indirectly through more global effects.

As a second approach to define the regions of the CRD that are involved in Wnt binding, we identified discrete surfaces on the CRD structure, and solvent-exposed residues constituting these surfaces in the mFz8 CRD were mutated *en bloc* to alanine. The sites of 15 different sets of alanine scanning mutations and their effects on XWnt8–AP binding are indicated in Fig. 3, and their positions on the CRD surface are shown in Fig. 4b. Most alanine scanning mutants had no effect on XWnt8–AP binding (green underlined symbols in Fig. 3), two blocks of mutations eliminated XWnt8–AP binding (red boxed symbols in Fig. 3), and three blocks greatly reduced binding (orange boxed symbols in Fig. 3).

The third method for identifying regions of the mFz8 CRD involved in Wnt binding exploited the inability of mouse Frizzled 6 (mFz6) or the isolated mFz6 CRD to bind XWnt8–AP²¹. Eight clusters of residues from the mFz6 CRD that occupy surface positions of the CRD and differ substantially in sequence from the corresponding regions of mFz8 were substituted for the homologous residues in mFz8, and the resulting mFz8/mFz6 chimaeras were assayed for XWnt8–AP binding (Figs 3 and 4c). Two substitutions resulted in a loss of XWnt8 binding, and these occurred within regions identified by the alanine scanning mutations as important for XWnt8–AP binding. One mFz8/mFz6 substitution in the region of the α 4 helix that had no effect on binding largely overlapped an alanine scanning mutation with reduced binding.

The three mutagenesis strategies produced maps of the CRD surface regions involved in Wnt binding that are in good agreement

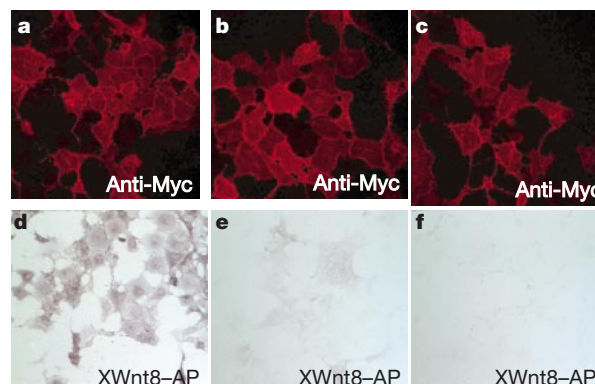


Figure 2 Binding of XWnt8–AP to mFz8 CRD–Myc–GPI displayed on the surface of transfected COS cells. COS cells were transiently transfected with an alanine scanning mutant encompassing either mFz8 residues 114–118, which bound XWnt8–AP (**a**, **d**), mFz8 residues 17–25, which showed reduced XWnt8–AP binding (**b**, **e**), or mFz8 residues 17–23, which did not bind XWnt8–AP (**c**, **f**). **a**–**c**, Immunostaining of cell-surface mFz8 CRD–Myc–GPI in living (that is, nonpermeabilized) cells with anti-Myc monoclonal antibody followed by Texas red-conjugated secondary antibody. Confocal image at 2 μ m thickness. **d**–**f**, XWnt8–AP binding as seen with alkaline phosphatase histochemistry.

and that strongly implicate a single region of the CRD surface, comprising three segments of the primary sequence, as important for Wnt binding (Fig. 4). The simplest interpretation of these experiments is that this surface is the site of direct contact between Wnt and CRD molecules. The implicated region of the CRD surface (950 Å²) could reasonably contact a single Wnt molecule ($M_r \sim 45K$). As a portion of the surface important for Wnt binding overlaps with the interface of the CRD dimer observed in the crystal, an alternative mechanism is suggested in which a subset of mutations may interfere with Wnt binding by hindering CRD dimerization.

Although no current evidence implicates dimerization of Wnt or

Frizzled proteins, the presence of the same dimer interface in the crystals of both sFRP-3 and mFz8 CRDs suggests that CRD dimerization may be of biological significance. Specifically, weak dimerization affinities of isolated CRDs may reflect dimerization induced by ligands and/or co-receptors as a feature of the signalling mechanism. As one test of this possibility, a solid-phase assay was developed that measures the XWnt8-dependent association of two differentially tagged mFz8 CRDs. In this assay, incubation with XWnt8-containing conditioned medium led to a >90-fold increase

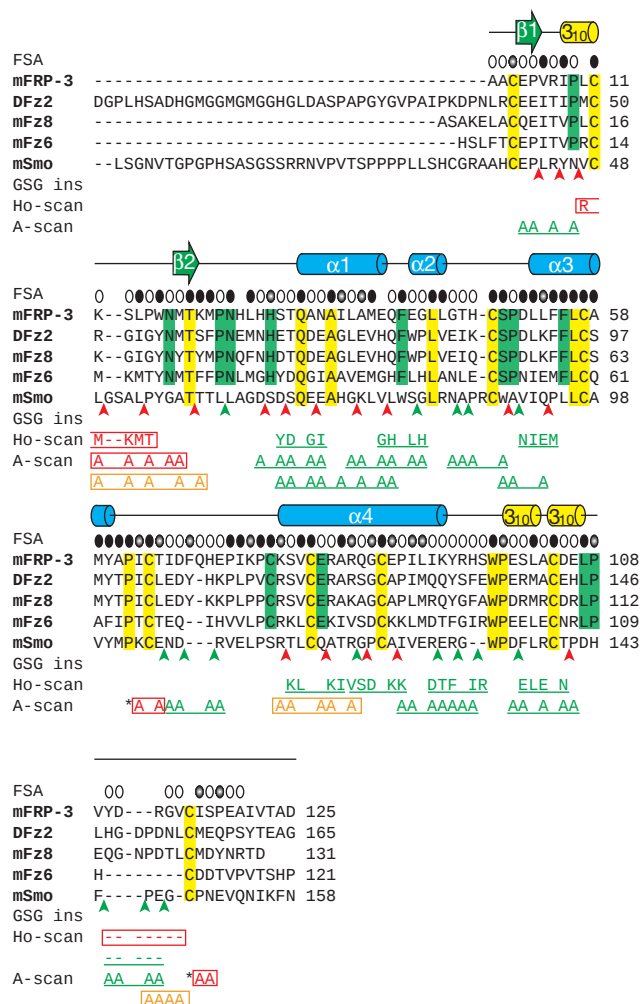


Figure 3 Structure-based alignment of CRD sequences indicating sites altered by mutagenesis. Residues conserved in the Frizzled and Smoothed families (yellow) or only in the Frizzled family (green) are indicated. Fractional solvent accessibility (FSA) is indicated above each residue for which sFRP-3 and mFz8 have similar accessibility: filled ovals for buried residues (<10% exposed); partially filled ovals for moderately exposed residues (10–40% exposed); open ovals for exposed residues (>40% exposed). GSG ins, sites of Gly-Ser-Gly tripeptide insertions (DFz2 CRD): red arrowheads, insertion sites that eliminate XWnt8–AP binding; green arrowheads, sites that do not affect XWnt8–AP binding. Ho-scan, sites of homologue scanning mutagenesis (mFz6 CRD into mFz8 CRD); A-scan, alanine scanning mutagenesis (mFz8 CRD): residues mutated *en bloc* are indicated by an underline or a box; mutations that do not affect (green), that diminish (orange) or that abolish (red) XWnt8–AP binding are indicated. mSmo, mouse Smoothed. Asterisk, two blocks adjacent in the tertiary structure and therefore constructed together as a single alanine mutation set. Sequences of the mature proteins were deduced from known signal-peptide cleavage sites and sites predicted by the program SignalP²⁹.

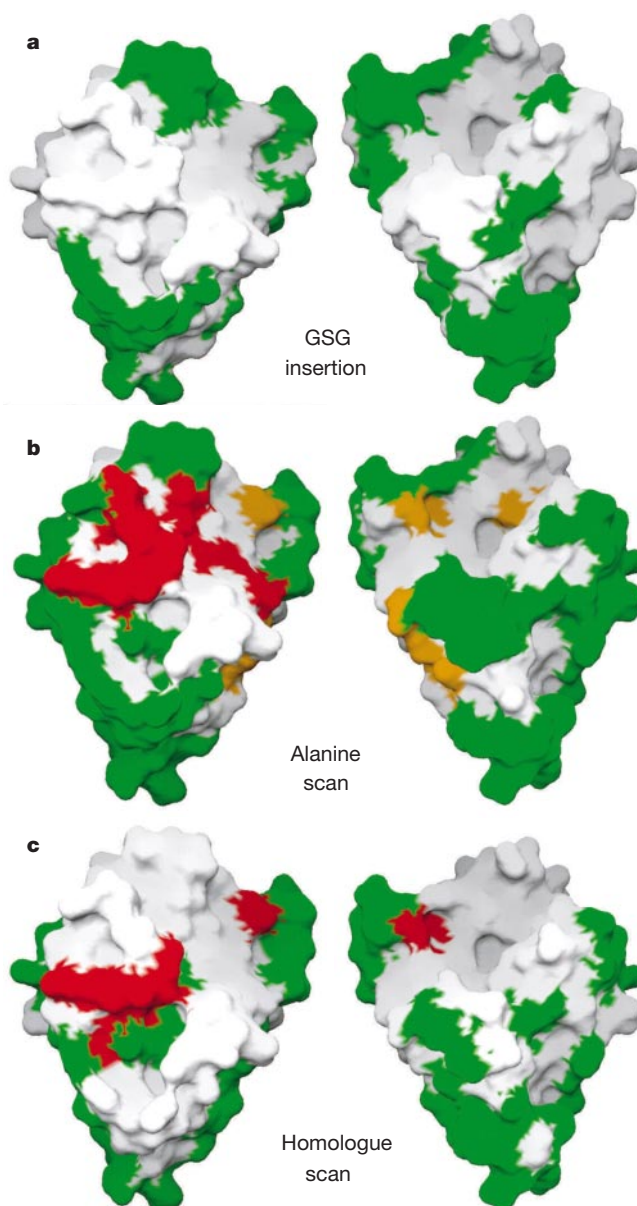


Figure 4 Surfaces involved in Wnt–CRD interactions modelled on the mFz8 CRD structure. Mutations that do not affect binding (green), that produce weak binding (orange) and that disrupt binding (red) are indicated. The orientation of the CRD surface on the left is identical to the ribbon diagrams in Fig. 1b; the orientation on the right is rotated 180° about a vertical axis. **a**, Locations of the 16 Gly-Ser-Gly (GSG) insertions in the DFz2 CRD that did not affect binding to XWnt8–AP (green indicates the two amino acids that flank the point of insertion). **b**, Alanine scanning mutations in the mFz8 CRD. Alanine substitution of residues 117 and 118 (green) in the mFz8 CRD permitted binding when associated with two alanine substitutions upstream of this location, but not when associated with two alanine substitutions downstream. **c**, Homologue scanning mutations for mFz6/mFz8 CRD. Deletion of residues 114–120 in the mFz8 CRD eliminates binding, but deletion of residues 114–118 (green) does not. Images were generated by GRASP³⁰.

in association of the tagged CRDs compared with control conditioned medium. The interpretation of this observation is uncertain, however, because the stoichiometry and composition of the CRD–Wnt complex formed in the assay is unknown. The majority of the XWnt8 present in conditioned medium is found in a large complex ($M_r > 600K$) as judged by gel filtration. Although these results are consistent with the possibility that Wnt-induced multimerization of Frizzled receptors is involved in Wnt signalling, a definitive test of this hypothesis must await the development of biochemically defined preparations of Wnt proteins.

The crystal structures of the CRDs reported here are the first for any member of this large class of homologous domains found in many important signalling molecules^{12,13} and appear to define a new protein fold. We used these structures to identify a surface on the CRD that is important for Wnt binding and we observed a dimer of both sFRP-3 and mFz8 CRDs that may be relevant to the mechanism of Wnt binding and signalling. These structures will allow a molecular understanding of Wnt–CRD specificity as more information concerning biological Wnt–CRD pairings becomes available, and will provide a framework for models of Wnt–CRD signal transduction. Finally, as the conserved cysteines permit accurate alignment of highly divergent CRD sequences (Fig. 3), the CRD structures provide a reliable model for studies of distantly related CRDs, such as those found in muscle-specific kinase and Smoothed^{12,13}, the functional roles of which are not known. □

Methods

Protein expression, purification and crystallization

Full-length complementary DNA from mouse sFRP-3 was subcloned into the pCIS vector and transfected into CHO cells using Lipofectin (Qiagen), and sFRP-3 protein was purified from conditioned serum-free medium (Ex-Cell 301, JRH Biosciences) by Mono-S cation exchange and heparin affinity chromatography (Pharmacia). Treatment of sFRP-3 with subtilisin results in a stable fragment that no longer binds heparin. N-terminal sequencing and mass spectrometry revealed this fragment to be the CRD. Full-length sFRP-3 did not crystallize, but the sFRP-3 CRD crystallized following enzymatic deglycosylation. The diffraction limit of these crystals depended on the level of deglycosylation but never extended beyond 7 Å. Asn 17 of sFRP-3 CRD (residues 1–125) was therefore mutated to Glu to eliminate the single N-linked glycosylation site; this mutant (sFRP-3 CRDM) was expressed in Chinese hamster ovary cells as a fusion protein with human growth hormone, an octahistidine tag, and a cleavage site for tobacco etch virus (TEV) protease¹⁴. This fusion protein was produced at 2.1 mg l⁻¹ and was initially purified by batch Ni-NTA immobilized metal affinity chromatography (Qiagen). The fusion protein was then cleaved by TEV protease, which leaves an additional Gly-Ser at the N terminus of the sFRP-3 CRDM. sFRP-3 CRDM was purified by gel filtration and Resource Q anion exchange chromatography (Pharmacia) before concentration to 15 mg ml⁻¹ in deionized H₂O for crystallization trials. mFz8 CRDM protein was expressed and purified in the same manner as sFRP-3 CRDM with the mutations N22E and N125E at putative glycosylation sites. All crystals were grown by vapour diffusion from hanging drops. For sFRP-3 CRDM, equal amounts of protein and a solution of 0.1 M HEPES and 12% PEG 3350 at pH 6.6 were mixed, suspended over this PEG solution, and equilibrated at 20 °C. sFRP-3 CRDM crystallized in space group P₂₁ with cell dimensions $a = 96.05$ Å, $b = 46.79$ Å, $c = 83.31$ Å and $\beta = 90.8^\circ$. mFz8 CRDM crystals were grown using 0.1 M HEPES and 1.9 M (NH₄)₂SO₄ at pH 7.3 as mother liquor. mFz8 CRDM crystallized in space group P₂ with cell dimensions $a = 55.69$ Å, $b = 36.37$ Å, $c = 57.94$ Å and $\beta = 98.82^\circ$.

Structure determination

All data were collected at beamline X4A of the National Synchrotron Light Source at Brookhaven National Laboratory. To introduce bromide ions for phasing¹⁶, sFRP-3 CRDM crystals were first transferred stepwise to buffer containing 0.1 M HEPES and 33% PEG 3350 at pH 6.6 and then transferred to the same buffer containing 0.5 M NaBr for 40 s before flash freezing in a nitrogen stream at 100 K. mFz8 CRDM crystals were flash frozen in a nitrogen stream after transfer to mineral oil. Diffraction data from crystals were collected at 100 K and processed with the program HKL²³. Nine solvent bromides were located and used for sFRP-3 CRDM phase determination with the program SOLVE²⁴. Phases were improved by solvent flattening and phase extension with the program DM²⁵. The program O was used for model building²⁶, and refinement was performed with the CNS package and maximum likelihood targets²⁷. Non-crystallographic symmetry restraints were used during early stages of the sFRP-3 CRDM refinement but relaxed when differences between molecules in the asymmetric unit became apparent. All refinements were carried out on data collected from a native crystal. Excluding the carboxy-terminal 5 residues, which extended away from the CRD and adopted multiple conformations, the six independent sFRP-3 CRD structures were essentially identical. The maximal pairwise root-mean-square deviation (RMSD) of alpha carbon distances between the six molecules in the asymmetric unit was 0.53 Å. The sFRP-3 CRDM with the lowest average B-factor and an ordered C terminus was chosen for display. mFz8 CRDM was solved by molecular

replacement using a sFRP-3 CRDM model in which all non-identical surface residues were truncated to alanine and the two structurally distinct loops removed. Of 122 alpha carbons of the mFz8 and sFRP-3 CRDs, 101 superimposed with a RMSD of 0.97 Å (Fig. 1b). Refinement and model building of mFz8 CRDM were carried out in the same manner as for the sFRP-3 CRDM without non-crystallographic symmetry restraints. The crystallographic R_{work}/R_{free} are 22.0/25.7 (30–1.9 Å) and 22.5/24.7 (30–1.35 Å), and the RMSD for bonds/angles are 0.0087 Å/1.48° and 0.0048 Å/1.29° for sFRP-3 and mFz8, respectively. Additional data collection and refinement statistics are available in the Supplementary Information.

Binding of XWnt8–AP to DFz2 and mFz8 CRD mutants

DFz2 CRD residues 1–248 and mFz8 CRD residues 1–146 were transiently expressed in COS cells as an N-terminal fusion to a Myc epitope tag and a GPI anchor. Tripeptide insertions (DFz2 CRD), alanine and homologue scanning mutants (mFz8 CRD) were expressed, and binding assays performed as described⁴. Briefly, CRD-transfected COS cells were incubated with conditioned medium containing XWnt8–AP, washed, fixed and stained with alkaline phosphatase substrates. Binding quality was assigned visually relative to native CRD binding to XWnt8–AP. Normal binding was equivalent to native, weak binding was any level less than native, and non-binding was undetectable activity of alkaline phosphatase.

Solid-phase assay of mFz8 CRD dimers

Plates with 96 wells (Corning) were coated with anti-hGH monoclonal antibody (ATCC) at 1 µg ml⁻¹ and blocked with 1% dialysed calf serum. Plates were then washed with PBS and 0.1% Tween-20 at pH 7.4 (PBS-T), and 3 µg of purified hGH-mFz8 CRD fusion protein was added. After incubation at 37 °C for 1 h, plates were washed with PBS-T, and 100 µl of either XWnt8–Myc or control S2 cell conditioned medium was added. After 15 min, 100 µl conditioned medium containing mFz8 CRD–AP expressed in transfected 293 cells was added, and the plates incubated for 45 min at 37 °C. Plates were then washed eight times with PBS-T, and the presence of mFz8 CRD–AP was detected using Blue Phos (Kirkegaard and Perry Laboratories).

Received 30 March; accepted 4 May 2001.

- Bhanot, P. *et al.* A new member of the frizzled family from *Drosophila* functions as a Wingless receptor. *Nature* **382**, 225–230 (1996).
- Nusse, R. & Varmus, H. E. Wnt genes. *Cell* **69**, 1073–1087 (1992).
- Wodarz, A. & Nusse, R. Mechanisms of Wnt signalling in development. *Annu. Rev. Cell Dev. Biol.* **14**, 59–88 (1998).
- Sparks, A. B., Morin, P. J., Vogelstein, B. & Kinzler, K. W. Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res.* **58**, 1130–1134 (1998).
- Rattner, A. *et al.* A family of secreted proteins contains homology to the cysteine-rich ligand-binding domain of frizzled receptors. *Proc. Natl Acad. Sci. USA* **94**, 2859–2863 (1997).
- Finch, P. W. *et al.* Purification and molecular cloning of a secreted, Frizzled-related antagonist of Wnt action. *Proc. Natl Acad. Sci. USA* **94**, 6770–6775 (1997).
- Leyns, L., Bouwmeester, T., Kim, S. H., Piccolo, S. & De Robertis, E. M. Frzb-1 is a secreted antagonist of Wnt signalling expressed in the Spemann organizer. *Cell* **88**, 747–756 (1997).
- Moon, R. T., Brown, J. D., Yang-Snyder, J. A. & Miller, J. R. Structurally related receptors and antagonists compete for secreted Wnt ligands. *Cell* **88**, 725–728 (1997).
- Wehrli, M. *et al.* *arrow* encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature* **407**, 527–530 (2000).
- Tamai, K. *et al.* LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000).
- Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).
- Masiakowski, P. & Yancopoulos, G. D. The Wnt receptor CRD domain is also found in MusK and related orphan receptor tyrosine kinases. *Curr. Biol.* **8**, R407 (1998).
- Xu, Y. K. & Nusse, R. The Frizzled CRD domain is conserved in diverse proteins including several receptor tyrosine kinases. *Curr. Biol.* **8**, R405–406 (1998).
- Leahy, D. J., Dann, C. E., Longo, P., Perman, B. & Ramyar, K. A mammalian expression vector for expression and purification of secreted proteins for structural studies. *Protein Expr. Purif.* **20**, 500–506 (2000).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Dauter, Z., Dauter, M. & Rajashankar, K. R. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr. D* **56**, 232–237 (2000).
- Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
- Orengo, C. A. *et al.* CATH—a hierarchic classification of protein domain structures. *Structure* **5**, 1093–1108 (1997).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Hsieh, J. C., Rattner, A., Smallwood, P. M. & Nathans, J. Biochemical characterization of Wnt–frizzled interactions using a soluble, biologically active vertebrate Wnt protein. *Proc. Natl Acad. Sci. USA* **96**, 3546–3551 (1999).
- Wells, J. A. Systematic mutational analyses of protein–protein interfaces. *Methods Enzymol.* **202**, 390–411 (1991).
- Otwiniowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).

25. Zhang, K. & Main, P. Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Crystallogr. A* **46**, 41–46 (1990).
26. Jones, T., Zou, J.-Y., Cowan, S. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
28. Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 961–985 (1991).
29. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6 (1997).
30. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Beneken, K. Ramyar, C. Ogata and the staff at X4A for help with data collection; Z. Dauter for advice and technical assistance with phase determination using solvent halides; and D. Andrew, P. Beachy, J. Berg, J. Taipale and W. Yang for comments on the manuscript. Structures reported are accession numbers 1IJX for sFRP-3 and 1IJY for mFzB. Supported by the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.J.L. (e-mail: leahy@groucho.med.jhmi.edu).

Mitochondrial endonuclease G is important for apoptosis in *C. elegans*

Jay Parrish*, Lily Li†, Kristina Klotz*, Duncan Ledwich*, Xiaodong Wang† & Ding Xue*

* Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado 80309, USA

† Howard Hughes Medical Institute and Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

Programmed cell death (apoptosis) is a tightly regulated process of cell disassembly in which dying cells and their nuclei shrink and fragment and the chromosomal DNA is degraded into internucleosomal repeats^{1–3}. Here we report the characterization of the *cps-6* gene, which appears to function downstream of, or in parallel to, the cell-death protease CED-3 of *Caenorhabditis elegans* in the DNA degradation process during apoptosis. *cps-6* encodes a homologue of human mitochondrial endonuclease G^{4,5}, and its protein product similarly localizes to mitochondria in *C. elegans*. Reduction of *cps-6* activity caused by a genetic mutation or RNA-mediated interference (RNAi) affects normal DNA degradation, as revealed by increased staining in a TUNEL assay, and results in delayed appearance of cell corpses during development in *C. elegans*. This observation provides *in vivo* evidence that the DNA degradation process is important for proper progression of apoptosis. CPS-6 is the first mitochondrial protein identified to be involved in programmed cell death in *C. elegans*, underscoring the conserved and important role of mitochondria in the execution of apoptosis.

Genetic analyses in *C. elegans* have led to the identification of several genes that are important for activation of programmed cell death (*egl-1*, *ced-3*, *ced-4* and *ced-9*), engulfment of cell corpses and degradation of DNA from the dead cells (*nuc-1*)⁶. However, little is known about what functions downstream of, or in parallel to, the activated CED-3 death protease to execute highly organized and efficient cell disassembly. To identify these missing links in the cell-death pathway, we carried out a sensitized genetic screen to isolate suppressors of a constitutively activated CED-3 mutant (acCED-3; D.L. and D.X., unpublished observations). In this screen, acCED-3 and green fluorescence protein (GFP) were co-expressed under the

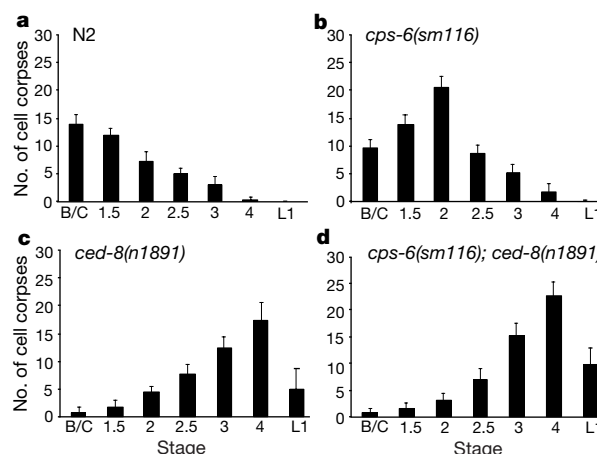


Figure 1 *cps-6(sm116)* causes delayed appearance of cell corpses during development. Cell corpses were scored in the following animals: **a**, N2; **b**, *cps-6(sm116)*; **c**, *ced-8(n1891)*; **d**, *cps-6(sm116); ced-8(n1891)*. Stages of embryos or larvae examined were: bean and comma (B/C), 1.5-fold (1.5), 2-fold (2), 2.5-fold (2.5), 3-fold (3), 4-fold (4) and early L1 larvae (L1). The definition of the stages of embryos is as described⁸. The y axis represents the average number of cell corpses scored in the head regions of embryos or larvae. Error bars are the standard error of the mean (s.e.m.). At least 15 embryos were counted for each developmental stage.

control of the promoters of the *mec-7* and *mec-3* genes (P_{mec-7} and P_{mec-3}), respectively, in six non-essential mechanosensory neurons⁷, and GFP was used as a sensitive cell-existence marker to select for mutants in which neuronal deaths induced by acCED-3 are delayed or partially suppressed. As shown in Table 1, expression of acCED-3 in animals homozygous for an integrated array (*smIs1*) containing both P_{mec-7} -acCED-3 and P_{mec-3} -GFP constructs induced most (>84%) of the six mechanosensory neurons to undergo apoptosis. We further sensitized this screen by introducing a mutation, *n1891*, into the *ced-8* gene of the *smIs1* animals, which results in a general delay of cell death⁸ and also partial suppression of acCED-3-induced neuronal deaths (Table 1). We then looked for mutations that could enhance the death-suppressing effect of *ced-8(n1891)*.

From one such screen of 3,000 mutagenized haploid genomes (see Methods), we identified a single recessive mutation, *sm116*, that significantly enhanced the suppression of acCED-3-mediated neuronal deaths by *ced-8(n1891)* (Table 1). *sm116* alone was also a weak suppressor. We found that *sm116* identifies a new locus on chromosome I, which was named *cps-6* (CED-3 protease suppressor).

To investigate whether *sm116* affects normal programmed cell death in nematodes, we performed a time-course analysis of embryonic cell corpses⁸. In wild-type animals, the peak of cell corpses appears in the bean or comma embryonic stage, while in the *ced-8(n1891)* mutant, the peak of cell corpses shifts to the four-fold embryonic stage, showing an overall delay in cell-corpse

Table 1 *cps-6(sm116)* enhances suppression of acCED-3-mediated neuronal deaths by *ced-8(n1891)*

Strain*	No. of animals	AVM†	ALMR/L†	PVM†	PLMR/L†
<i>smIs1</i>	50	16%	14%	10%	0%
<i>smIs1; ced-8(n1891)</i>	40	35%	8%	25%	0%
<i>cps-6(sm116); smIs1</i>	52	23%	24%	17%	1%
<i>cps-6(sm116); smIs1; ced-8(n1891)</i>	60	63%	35%	37%	3%

* *smIs1* was generated by integrating through exposure to γ-rays an extrachromosomal array containing P_{mec-7} -acCED-3 (10 μg ml⁻¹), P_{mec-3} -GFP (10 μg ml⁻¹) and pL15EK (40 μg ml⁻¹), a plasmid that rescues *lin-15* mutant⁹, into *lin-15(n765ts)* animals. It was then backcrossed six times with N2 animals and mapped to LGV.

† The percentage of mechanosensory neurons with GFP in L4 larvae is shown, which was scored using a fluorescent Nomarski microscope. AVM, ALMR/L, PVM and PLMR/L are six mechanosensory neurons.

appearance (Fig. 1a, c). In *cps-6(sm116)* animals, the peak of cell corpses shifts to the two-fold embryonic stage and the numbers of cell corpses in later embryonic stages also increase (Fig. 1b). These observations indicate that *sm116*, like the *ced-8* mutations, causes a delay in the appearance of cell corpses during development, an indication that the progression of apoptosis is delayed. Consistent with the result that *sm116* enhances the suppression of acCED-3 by

ced-8(n1891) (Table 1), we found that the appearance of embryonic cell corpses is further delayed in *cps-6(sm116); ced-8(n1891)* double mutants (Fig. 1d).

We then examined whether *cps-6(sm116)* prevents cell deaths and generates extra 'undead' cells in the anterior pharynx of mutant animals⁸. A strong loss-of-function mutation in *ced-3(n2433)* or *ced-4(n1162)* blocked nearly all cell deaths in *C. elegans* and resulted in a mean of 13.9 or 13.6 extra cells in the anterior pharynx, whereas in a weak *ced-3(n2447)* or *ced-4(n2273)* mutant, only a mean of 1.5 or 2.7 cells failed to die (Table 2). In *cps-6(sm116)* animals, a mean of 0.06 extra cells was observed (Table 2), indicating that *sm116* alone has an essentially undetectable effect in blocking apoptosis. In *cps-6(sm116); ced-8(n1891)* double mutants, a mean of 1.17 extra cells was seen, compared with 0.92 extra cells in *ced-8(n1891)* animals (Table 2). This is consistent with the observation above that these two mutations act together to suppress cell death. Furthermore, we found that *sm116* increases the number of extra cells observed in weak *ced-3(n2447)* or *ced-4(n2273)* mutants and in strong *ced-3(n2433)* or *ced-4(n1162)* mutants (Table 2). Together our results suggest that the *cps-6* gene is important for the normal progression of apoptosis and that its activity can further affect cell killing when the activity of essential cell-death components is compromised.

To understand how *cps-6* affects apoptosis, we cloned the *cps-6* gene. *cps-6(sm116)* was mapped to an interval between two cloned genes on chromosome I, *unc-11* and *dpy-5*, approximately 0.82 map units to the right of *unc-11* (Fig. 2a; see Methods). Transformation rescue experiments revealed that one cosmid in the mapped region, C41D11, which contains a homologue of mammalian endonuclease G (endoG) that has been implicated in mediating mammalian apoptotic DNA fragmentation⁹, fully rescues the *cps-6(sm116)* mutant (Fig. 2a). Additionally, we found that worm endoG generated a cell-death phenotype (RNAi) very similar to that of *sm116* (see below). Further deletion analyses of C41D11 indicate that only those C41D11 subclones containing an intact worm endoG (pJP9.8, pJP9.8(ΔBS), pJP5.5 and pJP5.3) but not the one (pJP9.8(ΔXba)) that lacks the first exon of worm endoG can rescue *cps-6(sm116)*, indicating that *cps-6* most likely corresponds to worm endoG

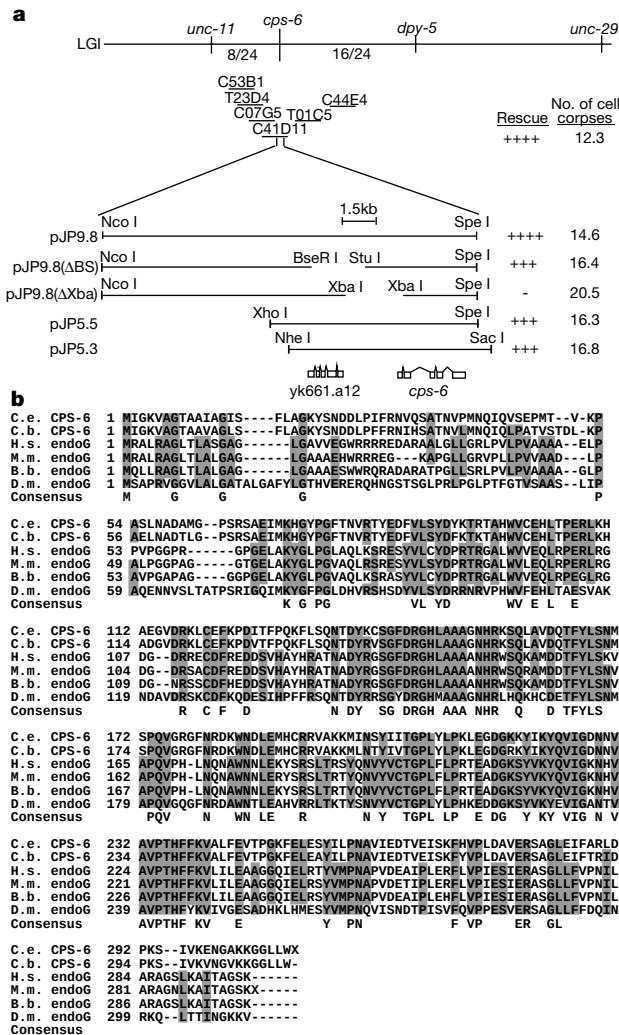


Figure 2 *cps-6* encodes an endoG homologue. **a**, Cloning of *cps-6*. The top bar represents the genetic map with the genes used for mapping *cps-6*. Numbers below are the fractions of the 24 recombinant events identified between *unc-11* and *dpy-5* that occur between *unc-11* and *cps-6* and between *cps-6* and *dpy-5*, respectively. The bottom panel shows *cps-6* rescue data obtained with a cosmid, C41D11, and various subclones of C41D11, as well as two ORFs identified in the region carried by pJP5.3. Subclones of C41D11 were generated using the indicated restriction enzymes. C41D11 or its subclone (10 μ g ml⁻¹ each) was injected into *cps-6(sm116); ced-8(n1891)* animals with *P_{sur-3}GFP* (20 μ g ml⁻¹) as a co-injection marker, which directs GFP expression in most cells throughout development²⁹. The number of cell corpses was scored in the head regions of four-fold fluorescent transgenic embryos. Four-fold embryos of *cps-6*; *ced-8* animals have a mean of 22.7 cell corpses, while *ced-8(n1891)* animals have a mean of 17.2 corpses. Transgenic lines that had a mean of <17.2 cell corpses in four-fold transgenic embryos were therefore considered as rescued. The numbers of cell corpses shown on the right are the mean numbers from all transgenic lines obtained for each construct (≥ 3 lines). **b**, Alignment of amino-acid sequences of the endoG proteins from *C. elegans* (C.e.), *C. briggsae* (C.b.), humans (H.s.), cattle (B.b.) and mice (M.m.), and the sequence from a protein (CG8862) predicted for *D. melanogaster* (D.m.). Residues that are identical in all six proteins are shaded and indicated as consensus residues. Residues that are identical in at least four proteins are also shaded.

Table 2 *cps-6* interacts with *ced-8*, *ced-3* and *ced-4* to affect programmed cell death

Strain*	No. scored	No. of extra cells†	
		Mean $\pm$ s.e.m.	Range
N2	20	0.00 $\pm$ 0.00	0
<i>ced-8(n1891)</i>	20	0.92 $\pm$ 0.79	0–3
<i>cps-6(sm116)</i>	18	0.06 $\pm$ 0.23	0–1
<i>cps-6(sm116); ced-8(n1891)</i>	18	1.17 $\pm$ 1.20	0–4
<i>ced-3(n2447)</i>	12	1.50 $\pm$ 0.67	1–3
<i>cps-6(sm116); ced-3(n2447)†</i>	18	3.17 $\pm$ 1.58	0–6
<i>ced-3(n2433)</i>	17	13.9 $\pm$ 1.22	12–16
<i>cps-6(sm116); ced-3(n2433)</i>	16	14.4 $\pm$ 1.67	11–16
<i>ced-4(n2273)</i>	22	2.68 $\pm$ 1.62	0–7
<i>cps-6(sm116); ced-4(n2273)†</i>	22	3.73 $\pm$ 1.58	1–6
<i>ced-4(n1162)</i>	12	13.6 $\pm$ 1.38	12–16
<i>cps-6(sm116); ced-4(n1162)†</i>	14	14.2 $\pm$ 1.25	12–16
N2; control (RNAi)	14	0.00 $\pm$ 0.00	0
N2; <i>cps-6(RNAi)</i>	20	0.20 $\pm$ 0.41	0–1
<i>ced-8(n1891)</i> ; control(RNAi)	18	0.89 $\pm$ 0.68	0–2
<i>ced-8(n1891); cps-6(RNAi)</i>	17	1.76 $\pm$ 0.83	1–3
<i>ced-3(n2447)</i> ; control(RNAi)	18	1.22 $\pm$ 1.26	0–4
<i>ced-3(n2447); cps-6(RNAi)</i>	22	3.14 $\pm$ 1.36	1–6
<i>ced-3(n2433)</i> ; control(RNAi)	12	13.8 $\pm$ 1.42	12–16
<i>ced-3(n2433); cps-6(RNAi)</i>	14	14.9 $\pm$ 1.17	13–16
<i>ced-4(n2273)</i> ; control(RNAi)	18	2.67 $\pm$ 1.57	1–6
<i>ced-4(n2273); cps-6(RNAi)</i>	18	4.28 $\pm$ 1.27	2–7
<i>ced-4(n1162)</i> ; control(RNAi)	12	13.5 $\pm$ 1.83	11–16
<i>ced-4(n1162); cps-6(RNAi)</i>	11	14.5 $\pm$ 1.75	11–16

* RNAi experiments were carried out as described in Methods. Control(RNAi) indicates that the animals were fed with bacteria containing an expression vector without an insert as a negative control.

† These strains also contain *dpy-5(e61)*.

‡ Extra cells were counted in the anterior pharynx of L4 hermaphrodites using Nomarski optics as described⁸.

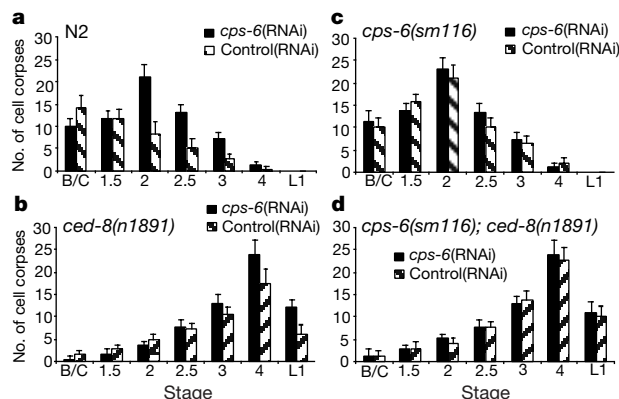


Figure 3 *Caenorhabditis elegans* endoG(RNAi) phenocopies *cps-6(sm116)*. L1 larvae from the following animals were fed with bacteria expressing worm *cps-6* double-stranded RNA (filled bars) or bacteria containing the expression vector without the *cps-6* insert (hatched bars) (see Methods): **a**, N2; **b**, *ced-8(n1891)*; **c**, *cps-6(sm116)*; **d**, *cps-6(sm116); ced-8(n1891)*. Progeny of RNAi-treated animals were counted for cell corpses at various developmental stages as described in Fig. 1 and the results were similarly depicted. At least 15 embryos or larvae were counted for each developmental stage.

(Fig. 2a; also see Supplementary Information). Finally, expression of worm endoG in *cps-6(sm116); ced-8(n1891)* embryos under the control of the promoter of the *dpy-30* gene (P_{dpy-30} -CPS-6), which is ubiquitously expressed in *C. elegans*¹⁰, fully rescues the Cps-6 phenotype (Table 3), providing further evidence that *cps-6* encodes worm endoG.

A BLAST (version 2.0) search revealed that worm endoG shares 48% identity and 69% similarity with human and mouse endoG. It also has significant sequence similarity to a predicted protein of *Drosophila melanogaster* (Fig. 2b). The high degree of sequence similarity among these proteins suggests that they are likely to have similar biochemical activities and functions.

When we used RNAi¹¹ to abolish the expression of worm endoG (see Supplementary Information), we found that endoG(RNAi) phenocopies the *sm116* mutation. For example, endoG(RNAi) similarly shifts the peak of cell corpses from the comma to the two-fold stage in N2 embryos (Fig. 3a). The cell-corpse profile of endoG(RNAi); *ced-8(n1891)* embryos is also very similar to that of *cps-6(sm116); ced-8(n1891)* embryos (Figs 1d and 3b). However, endoG(RNAi) does not significantly enhance the late-appearing cell-corpse phenotype associated with *cps-6(sm116)* or *cps-6(sm116); ced-8(n1891)* embryos (Fig. 3c, d), but it does so with another cell-death mutant, *cps-5* (data not shown). Furthermore, endoG(RNAi) also phenocopies *sm116* in suppressing acCED-3-

mediated ectopic deaths (see Supplementary Information). These results together confirm that *cps-6* encodes *C. elegans* endoG.

During the above studies, we found that *cps-6(RNAi)* animals contain a slightly higher mean number of extra cells (0.2 extra cells) in their anterior pharynx than *cps-6(sm116)* animals, indicating that *cps-6(RNAi)* animals might have a more severe cell-death phenotype and that *sm116* may not be a null allele (Table 2). Indeed, a higher mean number of extra cells was observed in *cps-6(RNAi); ced-8(n1891)* animals (1.76 extra cells) than in *cps-6(sm116); ced-8(n1891)* animals (1.17 extra cells; Table 2). *cps-6(RNAi)* also has stronger effects in enhancing the cell-survival phenotypes of both weak and strong *ced-3* and *ced-4* mutants than *sm116* (Table 2). These results provide further evidence that *cps-6* is important for proper progression or execution of apoptosis.

In mammals endoG is synthesized as a propeptide in the cytoplasm and imported into mitochondria through a process mediated by its amino-terminal mitochondrion-targeting sequence^{4,5}. This signal peptide is cleaved off upon entering mitochondria and the mature endoG can be released from mitochondria during apoptosis⁹. To determine the localization of CPS-6, we expressed a CPS-6::GFP fusion protein that contains the entire *cps-6* open reading frame (ORF) under the control of the *dpy-30* promoter (P_{dpy-30} -CPS-6::GFP) and found that it can fully rescue the Cps-6 phenotype (data not shown), indicating that this CPS-6::GFP fusion is targeted to its normal functional site. The punctate cytoplasmic staining pattern of CPS-6::GFP (Fig. 4b) coincides with that of Mitotracker Red (Fig. 4c, d), a fluorescent dye that specifically labels mitochondria¹², indicating that CPS-6 localizes to the mitochondria in *C. elegans*. Interestingly, when we expressed a

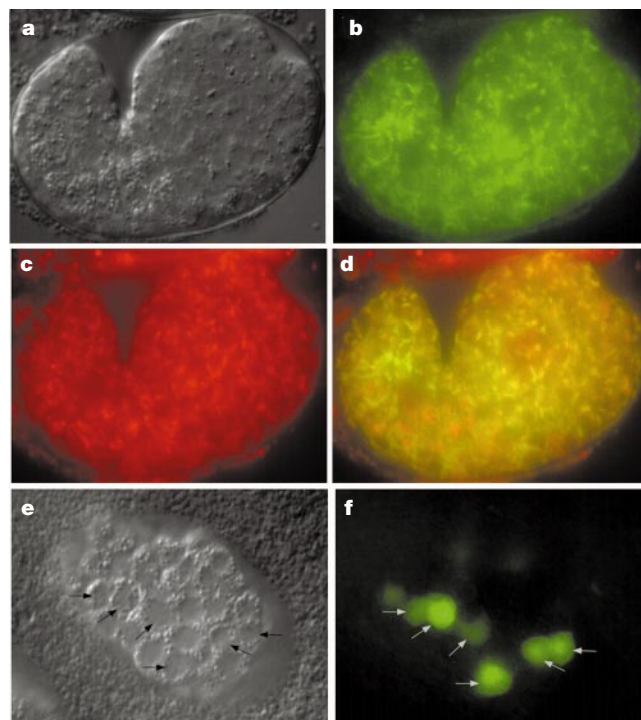


Figure 4 CPS-6 localizes to mitochondria in *C. elegans*. The following *C. elegans* embryos were examined with fluorescent microscopy to determine the localization of the fusion proteins: **a**, **b**, Nomarski (**a**) and fluorescent (**b**) images of an N2 embryo transgenic for the P_{dpy-30} -CPS-6::GFP construct. **c**, The staining pattern of Mitotracker Red (Molecular Probes) in the embryo shown in **b**. **d**, The merged image of **b** and **c**. **e**, **f**, Nomarski (**e**) and fluorescent (**f**) images of an N2 embryo transgenic for the P_{hsp} -CPS-6(22-308)::GFP constructs after the heat-shock treatment. Arrows indicate those nuclei that contain CPS-6(22-308)::GFP. Not all nuclei express GFP owing to the mosaic nature of the transgenic animals.

Table 3 Mouse endoG can substitute for *C. elegans* *cps-6*

Strain*	Array	No. of animals	No. of cell corpses†
<i>ced-8(n1891)</i>		22	17.3 ± 3.3
<i>cps-6(sm116); ced-8(n1891)</i>		15	22.7 ± 2.6
<i>cps-6(sm116); ced-8(n1891); P_{sur-5}GFP</i>	1	15	22.6 ± 3.0
	2	15	21.5 ± 2.8
	3	15	21.9 ± 3.6
<i>cps-6(sm116); ced-8(n1891); P_{dpy-30}CPS-6</i>	1	15	15.0 ± 4.2
	2	15	14.5 ± 3.7
	3	17	15.4 ± 3.5
<i>cps-6(sm116); ced-8(n1891); P_{dpy-30}mendoG</i>	1	16	16.9 ± 3.1
	2	16	17.1 ± 3.2
	3	15	17.5 ± 3.3
	4	15	16.7 ± 3.3

* P_{sur-5} GFP (20 μ g ml⁻¹) was injected into *cps-6(sm116); ced-8(n1891)* animals either alone as a control, or with P_{dpy-30} CPS-6 (25 μ g ml⁻¹) or P_{dpy-30} mendoG (25 μ g ml⁻¹) as a co-injection marker. Each numbered array represents an independent transgenic line.

†Cell corpses were scored in four-fold fluorescent transgenic embryos using Nomarski optics. Data are mean ± s.e.m.

truncated version of CPS-6::GFP without its putative mitochondrion-targeting sequence under the control of *C. elegans* heat-shock promoters (P_{hsp} CPS-6(22-308)::GFP), we found that upon heat-shock treatment this GFP-fusion protein is mainly localized in nuclei (Fig. 4e, f), indicating that the mature CPS-6, once released from mitochondria by appropriate signals, will probably go into nuclei to exert its function.

Given the sequence similarity between CPS-6 and endoG, we tested whether CPS-6 behaves as a nuclease *in vitro* and may function by cleaving genomic DNA of dying cells. When we incubated purified recombinant CPS-6 with isolated HeLa cell nuclei, we found that CPS-6 could cleave genomic DNA into nucleosomal fragments (Fig. 5a), a biochemical hallmark of apoptosis³. Characterization of the salt requirements of CPS-6 revealed that CPS-6 is dependent on magnesium (optimal concentration 1 mM; lane 4 of Fig. 5b), but not calcium, a biochemical property shared by human endoG (data not shown).

We then used the TUNEL (TdT-mediated dUTP nick end labelling) assay to examine whether CPS-6 affects DNA degradation in *C. elegans* apoptosis¹³. As shown before¹³, embryos from wild-type animals have very few TUNEL-positive cells (Fig. 6a). In contrast, *cps-6(sm116)* or *cps-6(RNAi)* embryos contain significantly higher numbers of TUNEL-positive cells, with most TUNEL-positive staining in embryos at the comma or 1.5-fold stage (Fig. 6b, c). This result indicates that CPS-6 is important in resolving TUNEL-positive DNA breaks generated during apoptosis. Consistent with this hypothesis, we found that very few TUNEL-positive cells were detected in *cps-6(sm116); ced-3(n2433)* or *cps-6(sm116); ced-9(n1950gf)* embryos (Fig. 6d, e), in which apoptosis is blocked by the *ced-3(n2433)* or *ced-9(n1950)* mutation, indicating that most TUNEL-positive cells seen in *cps-6(sm116)* embryos are probably from cells undergoing apoptosis. These TUNEL results strongly indicate that CPS-6 is involved in mediating apoptotic DNA fragmentation *in vivo*.

Several nucleases have previously been implicated in mediating

apoptotic DNA degradation³, including DFF40/CAD (40K DNA fragmentation factor or caspase-activated deoxyribonuclease), which is activated by caspases during apoptosis in mammals^{14,15}. In *C. elegans*, no DFF40/CAD homologue has been found but the *nuc-1* gene, which encodes a DNase II homologue, appears to affect DNA degradation during apoptosis^{13,16,17}. In *nuc-1(e1392)* mutant embryos¹³, means of 30–35 TUNEL-positive cells were observed (Fig. 6f), indicating a similar defect in resolving TUNEL-positive DNA breaks. However, unlike *cps-6*, *nuc-1* appears to be dispensable for apoptosis: cell-death activation and engulfment of cell corpses appear to proceed normally in *nuc-1* mutants^{13,16,17} (see Supplementary Information). Similarly, in mice lacking DFF40/CAD activity, although DNA fragmentation in apoptotic cells is significantly reduced, no evidence was found that apoptosis is either blocked or delayed^{18,19}. Thus our observation that loss of *cps-6* activity delays the progression of apoptosis provides the first genetic evidence that DNA degradation is an important component of cell-death execution.

Intriguingly, in *cps-6(sm116); nuc-1(e1392)* and *cps-6(RNAi); nuc-1(e1392)* embryos, the numbers of TUNEL-positive cells are significantly higher than those of single-mutant embryos, indicating that CPS-6 and NUC-1 may play partially redundant roles (Fig. 6g, h). However, we did not observe significant genetic interaction between *cps-6* and *nuc-1* in affecting the timing or activation of apoptosis (see Supplementary Information). This is consistent with the observed difference between the two nucleases in affecting apoptosis *in vivo*. CPS-6 and NUC-1 differ significantly in one other *in vivo* function: NUC-1, an endonuclease independent of Ca^{2+} and Mg^{2+} , is also required for degrading bacterial DNA in the

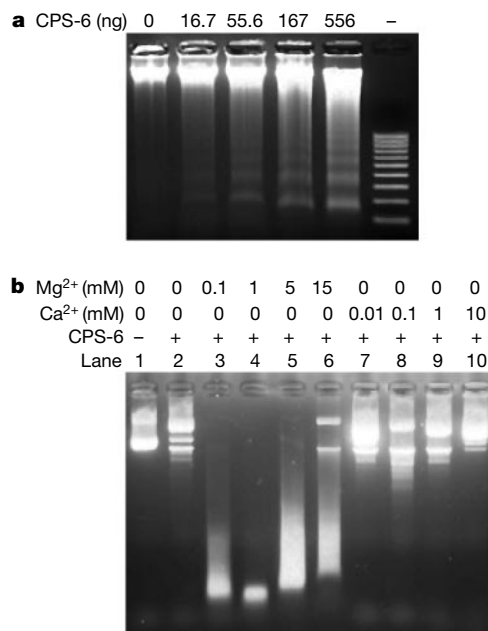


Figure 5 CPS-6 is a functional endonuclease *in vitro*. **a**, CPS-6 induces DNA laddering in isolated nuclei. Increasing amounts of purified recombinant CPS-6 were added to isolated HeLa cell nuclei and the nuclear DNA was extracted and resolved on the agarose gel. **b**, CPS-6 is a Mg^{2+} -dependent endonuclease. Purified CPS-6 (150 ng) was combined with 1 μ g of plasmid DNA in a variety of ionic conditions, and the DNA was resolved on the agarose gel.

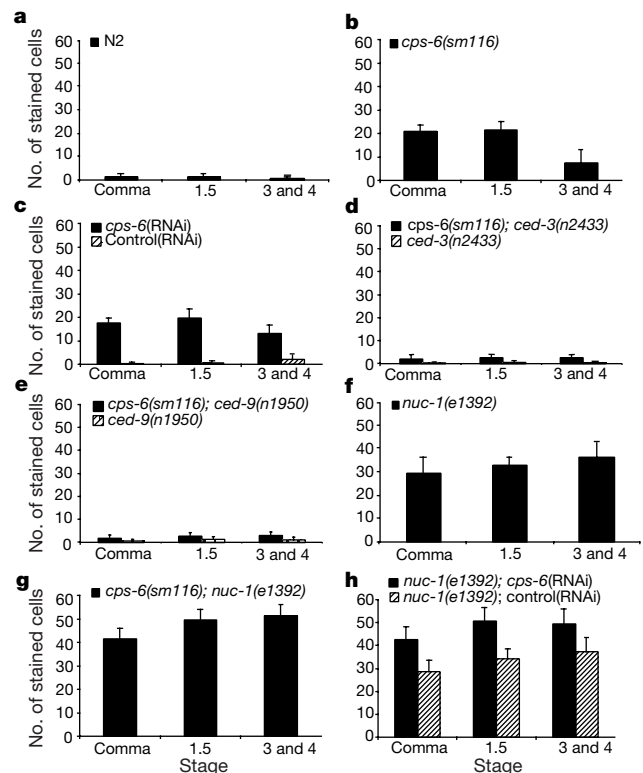


Figure 6 TUNEL staining in *cps-6* and *nuc-1* mutants. The following animals were stained with TUNEL: **a**, N2; **b**, *cps-6(sm116)*; **c**, N2; *cps-6(RNAi)* and N2; control(RNAi); **d**, *cps-6(sm116); ced-3(n2433)* and *ced-3(n2433)*; **e**, *cps-6(sm116); ced-9(n1950)* and *ced-9(n1950)*; **f**, *nuc-1(e1392)*; **g**, *cps-6(sm116); nuc-1(e1392)*; **h**, *nuc-1(e1392); cps-6(RNAi)* and *nuc-1(e1392); control(RNAi)*. The stages of embryos examined were: comma, 1.5-fold (1.5) and 3- and 4-fold (3 and 4). The y axis represents the mean numbers of TUNEL-positive cells present in embryos scored. Error bars are s.e.m. At least 15 embryos were scored for each developmental stage.

intestinal lumen, a function not associated with CPS-6 (data not shown). CPS-6 thus appears to be a cell-death-specific nuclease.

To test whether CPS-6 and endoG are functional homologues, we expressed mouse endoG (*P_{dpv-30}mendoG*) in *cps-6(sm116); ced-8(n1891)* animals and found that mendoG can fully rescue the *Cps-6* phenotype (Table 3). The fact that mendoG can replace the function of *cps-6* in nematodes and the findings that CPS-6 and mammalian endoG are strikingly similar in biochemical properties, subcellular localization and their involvement in apoptotic DNA degradation⁹, strongly indicate that CPS-6 and mammalian endoG define an evolutionarily conserved family of nucleases that are important in mediating apoptotic DNA degradation.

One major unresolved issue is the role of mitochondria in apoptosis. Overwhelming evidence suggests that mitochondria play a critical role in mammalian apoptosis²⁰. This is supported by the identification of crucial apoptogenic factors such as cytochrome *c*, Smac (also known as DIABLO) and AIF (apoptosis-inducing factor), which are released from mitochondria to activate different aspects of apoptosis^{21–24}, and the findings that many Bcl-2 family proteins, key regulators of apoptosis, are localized to, or capable of, targeting to mitochondria²⁰. However, there is no definitive evidence so far for the involvement of mitochondria in invertebrate apoptosis. In our study, we demonstrated that CPS-6, a mitochondrial protein of *C. elegans*, is involved in the nuclear DNA-fragmentation process of apoptosis, in much the same way as mammalian endoG. These studies provide the most convincing evidence to date that mitochondria are involved in apoptosis in an invertebrate and that the role of mitochondria as important regulators of programmed cell death is evolutionarily conserved. Thus, the DNA-degradation pathway involving CPS-6/endoG represents an important ancient link between the mitochondria and apoptosis. □

Methods

Strains and culture conditions

Strains of *C. elegans* were maintained using standard procedures²⁵. Most of the alleles used in this study are described in ref. 26, except two *ced-3* alleles, *n2447* and *n2433* (ref. 27).

Mutagenesis

Standard ethylmethane sulphonate (EMS) mutagenesis²⁵ was performed on *smIs1; ced-8(n1891)* hermaphrodites. F₂ progeny were screened for mutants in which GFP was seen in multiple mechanosensory neurons, using a fluorescent dissecting microscope. F₃ self progeny of candidate mutants that maintained the phenotype were then selected, backcrossed with *ced-8(n1891)* animals several times, and analysed further.

Genetic mapping

Standard genetic techniques²⁵ were used to map *cps-6(sm116)* to LGI using its *Cps* phenotype (increasing the number of cell corpses in *ced-8(n1891)* four-fold embryos). We used three-factor mapping to locate the position of *cps-6* relative to *dpv-5* and *unc-11* or *unc-29*. We first mapped *cps-6* to the left or right of *dpv-5*; from *dpv-5 unc-29/cps-6; ced-8(n1891)* mothers, 12 of 12 Unc non Dpy lines contained *cps-6(sm116)*, whereas 0 of 10 Dpy non Unc lines contained *cps-6(sm116)*. We next mapped *cps-6* between *unc-11* and *dpv-5*; from *unc-11 dpv-5/cps-6; ced-8(n1891)* mothers, 16 of 24 Dpy non Unc lines contained *cps-6(sm116)*.

Counting of cell corpses and extra cells

The number of cell corpses in the head regions of living embryos or L1-stage larvae and the number of extra cells in the anterior pharynx of L4-stage animals were determined with Nomarski optics as described⁸.

Molecular biology

Standard methods of polymerase chain reaction amplification, cloning and sequencing were used (see Supplementary Information).

RNAi experiments

RNAi experiments were carried out using a previously described feeding method²⁸, with a minor modification that we transferred ~10 L1-stage hermaphrodites to RNAi plates and scored the progeny of these animals for cell-death phenotypes.

TUNEL assays

The TUNEL assays were carried out as described¹³, using an *in situ* cell-death detection kit (Roche).

Recombinant CPS-6 and the nuclease assays

Full-length *cps-6* complementary DNA with an additional carboxy-terminal six-histidine-tag was cloned into the baculovirus expression vector pFastBac1 (Life Technologies). The expression and purification of the recombinant His-tagged CPS-6 were carried out as described⁹. The nuclease assays were also carried out as described⁹ using purified recombinant CPS-6.

Received 20 February; accepted 29 May 2001.

- Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239–257 (1972).
- Steller, H. Mechanisms and genes of cellular suicide. *Science* **267**, 1445–1449 (1995).
- Nagata, S. Apoptotic DNA fragmentation. *Exp. Cell Res.* **256**, 12–18 (2000).
- Ruiz-Carrillo, A. & Renaud, J. Endonuclease G: a (dG)n × (dC)n-specific DNase from higher eukaryotes. *EMBO J.* **6**, 401–407 (1987).
- Cote, J. & Ruiz-Carrillo, A. Primers for mitochondrial DNA replication generated by endonuclease G. *Science* **261**, 765–769 (1993).
- Horvitz, H. R. Genetic control of programmed cell death in the nematode *Caenorhabditis elegans*. *Cancer Res. (Suppl.)* **59**, 1701–1706 (1999).
- Chalfie, M. The differentiation and function of the touch receptor neurons of *Caenorhabditis elegans*. *Prog. Brain Res.* **105**, 179–182 (1995).
- Stanfield, G. M. & Horvitz, H. R. The *ced-8* gene controls the timing of programmed cell deaths in *C. elegans*. *Mol. Cell* **5**, 423–433 (2000).
- Li, L. Y., Luo, X. & Wang, X. Endonuclease G is an apoptotic DNase when released from mitochondria. *Nature*, **412**, 95–99 (2001).
- Hsu, D. R., Chuang, P. T. & Meyer, B. J. DPY-30, a nuclear protein essential early in embryogenesis for *Caenorhabditis elegans* dosage compensation. *Development* **121**, 3323–3334 (1995).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Chen, F. *et al.* Translocation of *C. elegans* CED-4 to nuclear membranes during programmed cell death. *Science* **287**, 1485–1489 (2000).
- Wu, Y. C., Stanfield, G. M. & Horvitz, H. R. NUC-1, a *Caenorhabditis elegans* DNase II homolog, functions in an intermediate step of DNA degradation during apoptosis. *Genes Dev.* **14**, 536–548 (2000).
- Liu, X., Zou, H., Slaughter, C. & Wang, X. DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
- Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).
- Sulston, J. E. Post-embryonic development in the ventral cord of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **275**, 287–297 (1976).
- Hedgecock, E. M., Sulston, J. E. & Thomson, J. N. Mutations affecting programmed cell deaths in the nematode *Caenorhabditis elegans*. *Science* **220**, 1277–1279 (1983).
- Zhang, J. *et al.* Resistance to DNA fragmentation and chromatin condensation in mice lacking the DNA fragmentation factor 45. *Proc. Natl Acad. Sci. USA* **95**, 12480–12485 (1998).
- McIlroy, D. *et al.* An auxiliary mode of apoptotic DNA fragmentation provided by phagocytes. *Genes Dev.* **14**, 549–558 (2000).
- Gross, A., McDonnell, J. M. & Korsmeyer, S. J. BCL-2 family members and the mitochondria in apoptosis. *Genes Dev.* **13**, 1899–1911 (1999).
- Liu, X., Kim, C. N., Yang, J., Jemerson, R. & Wang, X. Induction of apoptotic program in cell-free extracts: requirement for dATP and cytochrome *c*. *Cell* **86**, 147–157 (1996).
- Susin, S. A. *et al.* Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397**, 441–446 (1999).
- Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome *c*-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
- Verhagen, A. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing inhibitor of apoptosis (IAP) proteins. *Cell* **102**, 43–53 (2000).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Riddle, D. L., Blumenthal, T., Meyer, B. J. & Preiss, J. R. (eds) *C. elegans II* (Cold Spring Harbor Laboratory, Cold Spring Harbor, 1997).
- Shaham, S., Reddien, P. W., Davies, B. & Horvitz, H. R. Mutational analysis of the *Caenorhabditis elegans* cell-death gene *ced-3*. *Genetics* **153**, 1655–1671 (1999).
- Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
- Gu, T., Orita, S. & Han, M. *Caenorhabditis elegans* SUR-5, a novel but conserved protein, negatively regulates LET-60 Ras activity during vulval induction. *Mol. Cell Biol.* **18**, 4556–4564 (1998).
- Clark, S. G., Lu, X. & Horvitz, H. R. The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* **137**, 987–997 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank A. Coulson and the *C. elegans* Genome Sequencing Consortium for cosmid; A. Fire for vectors; B. Hersch and R. H. Horvitz for the protocol on using Mitotracker Red; and Y. C. Wu for discussion. We also thank R. J. McIntosh, G. Odorizzi, R. Poynton and W. Wood for comments on the manuscript. J.Z.P. is supported by a NSF Graduate Fellowship and X.W. by grants from NIH and the Welch Foundation. D.X. is supported by grants from ACS and NIH and is a recipient of the BWF Career Award and the Searle Scholar Award.

Correspondence and requests for materials should be addressed to D.X. (e-mail: ding.xue@colorado.edu). The sequence for *cps-6* is deposited in GenBank under accession number AF390558.

Endonuclease G is an apoptotic DNase when released from mitochondria

Lily Y. Li, Xu Luo & Xiaodong Wang

Howard Hughes Medical Institute & Department of Biochemistry,
University of Texas Southwestern Medical Center at Dallas, Dallas,
Texas 75390, USA

Nucleosomal fragmentation of DNA is a hallmark of apoptosis (programmed cell death)¹, and results from the activation of nucleases in cells undergoing apoptosis. One such nuclease, DNA fragmentation factor (DFF, a caspase-activated deoxyribonuclease (CAD) and its inhibitor (ICAD)), is capable of inducing DNA fragmentation and chromatin condensation after cleavage by caspase-3 (refs 2–4). However, although transgenic mice lacking DFF45 or its caspase cleavage site have significantly reduced DNA fragmentation^{5,6}, these mice still show residual DNA fragmentation and are phenotypically normal^{5–7}. Here we report the identification and characterization of another nuclease that is specifically activated by apoptotic stimuli and is able to induce nucleosomal fragmentation of DNA in fibroblast cells from embryonic mice lacking DFF. This nuclease is endonuclease G (endoG), a mitochondrion-specific nuclease that translocates to the nucleus during apoptosis. Once released from mitochondria, endoG cleaves chromatin DNA into nucleosomal fragments independently of caspases. Therefore, endoG represents a caspase-independent apoptotic pathway initiated from the mitochondria.

After the activation of cell-surface death receptors such as Fas/CD95, caspase-8 cleaves Bid, thus freeing the truncated Bid (tBid) to translocate to the mitochondria and release apoptogenic factors such as cytochrome *c*, which in turn triggers the activation of caspase-9 (refs 8–10). This reaction has been reconstituted *in vitro*

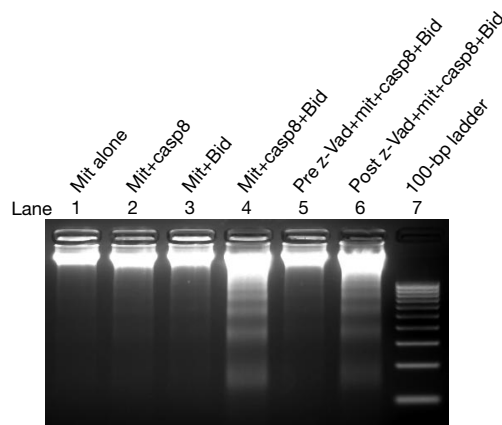


Figure 1 Identification of the mitochondrial nuclease. Isolated mice liver mitochondria (mit) were incubated with buffer alone (lane 1), with purified recombinant Bid or caspase-8 (casp8) separately (lanes 2, 3), or with both recombinant Bid and caspase-8 (lane 4). Reactions were prepared by either pre-incubating 20 μ M z-Vad-fmk (z-Vad) with caspase-8 for 15 min at 30 °C before adding Bid and purified mitochondria (lane 5), or pre-incubating with caspase-8 and Bid for 15 min at 30 °C then with 20 μ M z-Vad-fmk for 15 min at 30 °C before adding purified mitochondria (lane 6). The reactions were then incubated and centrifuged (see Methods) and the supernatants were collected. Aliquots of 38 μ l of the supernatant fluid were incubated with 17 μ l of buffer D (buffer A, 10 mM MgCl₂ and 100 μ M CaCl₂) and 5 μ l of purified nuclei for 2 h at 37 °C and purified (see Methods). Samples of the resuspended DNA were electrophoresed on a 1.2% agarose gel in 0.5 $\times$ TBE (45 mM Tris-HCl at pH 8.0, 45 mM boric acid, 1 mM EDTA) after RNase treatment.

with purified mitochondria, recombinant Bid and recombinant caspase-8. Using this system, we looked for additional apoptogenic factors, particularly DNase activities released from the mitochondria, by incubating the supernatant of Bid plus caspase-8 treated mitochondria with isolated nuclei from normal HeLa cells. As shown in Fig. 1, the nuclei incubated with the supernatant from untreated mitochondria (lane 1), or with the supernatant from mitochondria treated with either full-length Bid or caspase-8 alone (lanes 2 and 3) showed no detectable DNA fragmentation. In contrast, the supernatants from mitochondria incubated with both caspase-8 and Bid induced DNA fragmentation in the co-incubated nuclei (lane 4). The pre-incubation of a caspase inhibitor (z-Vad-fmk) with caspase-8 abolished the activity of the DNase released from the mitochondria (lane 5), indicating that the generation of tBid is necessary for the release of this activity. However, when we pre-incubated caspase-8 and Bid to generate tBid first, and then incubated the mixture with z-Vad-fmk before adding the mitochondria, the DNA-fragmentation activity was still present in the treated mitochondrial supernatant (lane 6). These results indicate that the release of this DNase from mitochondria is dependent on the generation of tBid, but that the activity of the DNase is independent of caspase once it is released from mitochondria.

Using DNA fragmentation in the co-incubated nuclei as an assay, we fractionated the supernatant of mouse liver mitochondria treated with recombinant Bid and caspase-8, and purified the DNase released from mitochondria with a six-step protocol (see Methods). The result of the last step of purification, a Superdex 200 gel-filtration chromatography column, is shown in Fig. 2. The DNA-fragmentation activity was observed in column fractions 11–13 (upper panel). The same fractions were also subjected to 15% SDS–polyacrylamide gel electrophoresis (PAGE) and the proteins were visualized by silver staining. A protein band from the silver-stained gel with a relative molecular mass of 30,000 (*M*_r 30K) correlated with the DNA-fragmentation activity (lower panel).

To obtain the sequence identity of this putative DNase, the protein was excised from the gel and subjected to tryptic digestion followed by mass-spectral analysis. The mass fingerprinting of the digested peptides matched that of a known mouse mitochondrial nuclease, endoG (see Supplementary Information). EndoG is encoded by a nuclear gene and translated in the cytosol with the amino-terminal 48 amino acids being the mitochondrial targeting

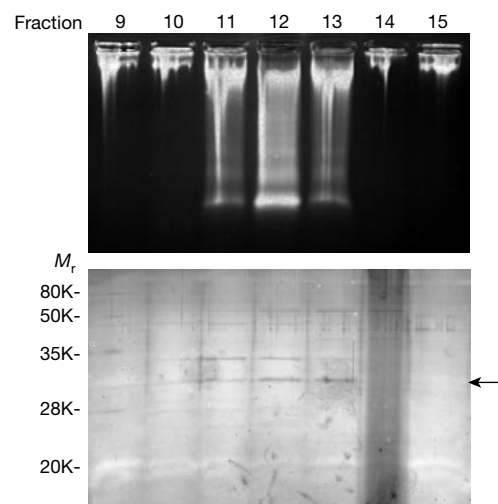


Figure 2 Purification of the mitochondrial nuclease (arrowed). Aliquots from the last purification step, the Superdex 200 column, were assayed for DNA fragmentation activity by incubation with isolated nuclei (upper panel; see Methods). The same samples were also subjected to 15% SDS–PAGE and silver stained (lower panel).

sequence, which is subsequently removed when the protein is imported into mitochondria^{11,12}.

To confirm that the DNase released from mitochondria by tBid is endoG, we generated a polyclonal antibody against the bacterially expressed endoG (amino acids 127–286) and performed western blotting and immunodepletion experiments with this antibody (Fig. 3a). This antibody specifically recognizes the precursor and the mature forms of both recombinant human endoG generated with a baculovirus expression system and murine endoG in mouse liver mitochondria (middle panel, lanes 1 and 2, respectively). The mature recombinant human endoG is slightly larger than mouse endoG possibly owing to the different processing in insect cells, the presence of a six-histidine tag, and additional amino acids. The pre-immune serum from the same animal failed to recognize these two bands (Fig. 3a, left panel). To further prove the specificity of the

antibody, we showed that the addition of recombinant endoG during western blot can block the immunoreactivity toward these two bands (right panel). The depletion of endoG from the tBid-treated mitochondrial supernatant with the anti-endoG antibody eliminated the DNA fragmentation (Fig. 3b, upper panel, lane 6) as well as the plasmid-cleaving DNase activity (lower panel, lane 6), indicating that endoG is the only DNase present in the tBid-treated mitochondrial supernatant. In contrast, the same depletion experiment performed with the pre-immune serum did not deplete the DNase activity (lane 5).

We next tested whether endoG alone is sufficient to cleave nuclear DNA into a discrete ladder using the purified recombinant endoG. As shown in Fig. 3c, increasing amounts of purified recombinant endoG induced increasing levels of DNA fragmentation in the co-incubated nuclei and cleaved plasmid DNA into smaller fragments. These results indicate that endoG is sufficient to generate DNA fragmentation.

We also compared the time course of release for both endoG and cytochrome *c* induced by purified tBid using a mitochondrial matrix protein, heat shock protein 70 (mtHsp70), as a control. As shown in Fig. 4, in the absence of tBid, neither cytochrome *c* nor endoG was released from mitochondria even after incubation for 90 min (lane 9). In the tBid-treated mitochondria, both endoG and cytochrome *c* were observed in the mitochondrial supernatants after 15 min of incubation (lane 4). By 90 min, most of the cytochrome *c* and endoG were released from mitochondria (lane 10). In contrast, the mtHsp70 was not released from mitochondria, indicating the inner membrane of mitochondria remained intact.

To test whether endoG is released in apoptotic cells, we performed immunofluorescent staining of both normal and apoptotic cells (Fig. 5). We chose mouse embryonic fibroblast (MEF) cells from the DFF45/ICAD-knockout mice for this study because these cells have lost the activities of caspase-3-dependent DNA fragmentation and chromatin condensation⁵. Immunostaining of untreated DFF45/ICAD-knockout MEF cells with the anti-endoG antibody revealed a

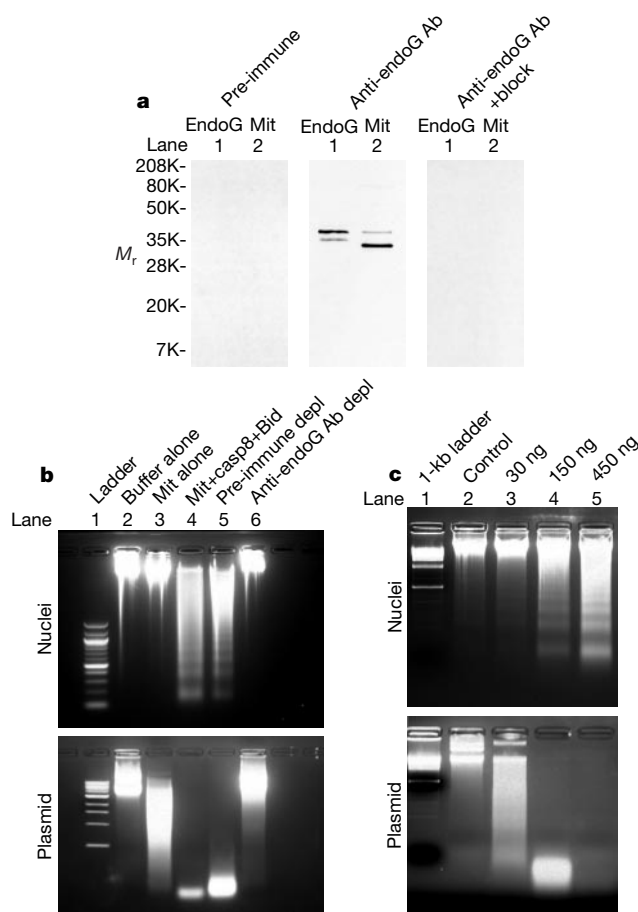


Figure 3 EndoG is both necessary and sufficient to cause DNA fragmentation. **a**, Recombinant human endoG and mouse mitochondrial pellet (Mit) were resuspended in 1× SDS loading buffer and subjected to 15% SDS–PAGE. The samples were transferred to a nitrocellulose membrane and western blots were performed using either pre-immune samples, anti-endoG antibodies (Ab) or endoG antibodies in the presence of recombinant endoG, as indicated. **b**, Isolated mice liver mitochondria were incubated with buffer alone or with recombinant Bid and caspase-8 (casp8). After centrifugation, aliquots of the supernatant from the treated mitochondria were immunodepleted (depl) with protein A-agarose beads coated either with control pre-immune antibodies or anti-endoG antibodies, and the DNA-fragmentation assay with isolated nuclei was performed. Resuspended DNA from the assays was visualized after electrophoresis on a 1.2% agarose gel in 0.5 × TBE gel. **c**, Recombinant human endoG was prepared (see Methods) and the protein concentrations were determined with the Bradford assay. Titration of the purified recombinant protein with both nuclear fragmentation and plasmid cleavage assays was performed to test the nuclease activity of the recombinant protein. The resulting samples were loaded onto a 1.2% agarose gel in 0.5 × TBE gel.

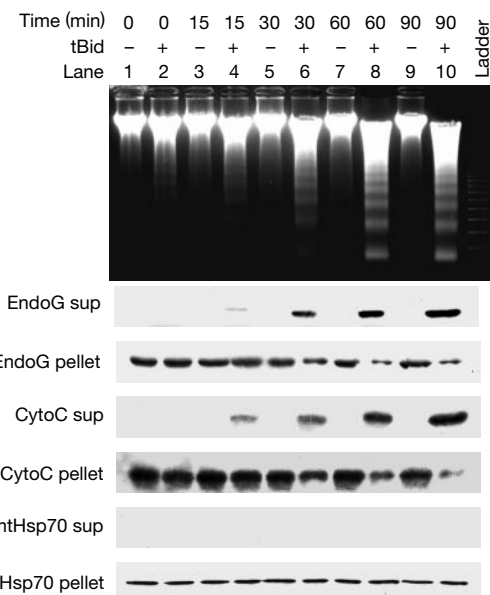


Figure 4 Characterization of the release process of endoG. Isolated mitochondria from mouse liver were incubated with either control buffer or purified recombinant tBid as described¹⁶ for up to 90 min at 30 °C. Aliquots were taken at each time point as indicated and subjected to centrifugation at 12,000g for 15 min. The resulting supernatants (sup) were used for DNA-fragmentation activity (upper panel) and Western blot analysis using antibodies against cytochrome *c* (CytoC), endoG, and mtHsp70. The resulting mitochondrial pellets were subjected to Western blot analysis using antibodies against these three proteins.

punctate mitochondrial pattern that co-localizes with cytochrome *c*. The staining of endoG but not cytochrome *c* was blocked by the addition of recombinant endoG. Upon induction of apoptosis by ultraviolet irradiation or treatment in TNF with cyclohexamide (CHX), both endoG and cytochrome *c* were released from the mitochondria to cytosol and nuclei.

The immunostaining data indicate that the DFF-knockout cells might still fragment their DNA, albeit at a reduced level when undergoing apoptosis, possibly owing to the release of endoG from mitochondria. To test this possibility, we analysed DNA fragmentation in the wild-type and DFF45-knockout MEF cells undergoing apoptosis induced by ultraviolet irradiation and TNF with CHX. As shown in Fig. 6a, both ultraviolet irradiation and TNF/CHX treatment induced DNA fragmentation in wild-type and DFF-knockout MEF cells, although the wild-type cells showed more DNA fragmentation (lanes 5, 6, 9 and 10). When we pre-incubated cells with media containing z-Vad-fmk before ultraviolet treatment, the level of DNA fragmentation decreased in the wild-type cells but was largely unaffected in the DFF-knockout cells (lanes 7 and 8). In contrast, z-Vad-fmk essentially eliminated the ability of TNF/CHX to induce DNA fragmentation in both the wild-type and the DFF-knockout MEF cells (lanes 11 and 12). This finding again confirmed that caspase-8 activation in TNF-induced apoptosis is upstream of the mitochondrial damage and the release of apoptogenic factors¹³. In contrast, caspase activation occurs after mitochondrial damage and cytochrome *c* release following ultraviolet irradiation¹⁴. Therefore the caspase inhibitor can inhibit DNA fragmentation caused by DFF, which has an absolute dependence on activated caspase for its activity, but can not inhibit the release of endoG from mitochondria following ultraviolet irradiation.

To confirm that the DNA-fragmentation activity observed in the DFF45-knockout MEF cells is caused by endoG, we prepared

extracts containing both nuclei and mitochondria from the DFF45-knockout MEF cells and induced the fragmentation of chromatin in the nuclei with tBid. As shown in Fig. 6b, addition of tBid induced nucleosomal fragmentation in the nuclei prepared from DFF45-knockout cells (lane 2). Pre-incubation of tBid with about twofold more of the death-preventing protein Bcl-x_L completely blocked the DNA fragmentation, indicating that the DNA fragmentation was caused by factors released from the co-incubated mitochondria and that the release process was regulated by Bcl-x_L (lane 3). Addition of the polyclonal antibody against endoG in the reaction also completely blocked DNA fragmentation (lane 5) while the pre-immune serum showed no effect (lane 4).

The persistence of DNA fragmentation induced by ultraviolet irradiation in the presence of caspase inhibitor indicated that DNA fragmentation could be induced by endoG independently of caspase. To further demonstrate this point, we used another pro-apoptotic protein, Bim, to repeat the experiment shown in Fig. 6b. Like Bid, Bim is a member of the 'BH3-only' pro-death family of proteins and can release cytochrome *c* *in vitro* as efficiently as tBid (data not shown). Unlike Bid, the apoptotic activity of Bim is regulated through its dissociation from microtubules during apoptosis in a caspase-independent fashion¹⁵. As shown in Fig. 6c, purified recombinant Bim fused to glutathione S-transferase (GST) induced DNA fragmentation in extracts prepared from DFF-knockout cells (lane 2) whereas GST protein had no effect (lane 6). As with tBid, both Bcl-x_L and the antibody against endoG

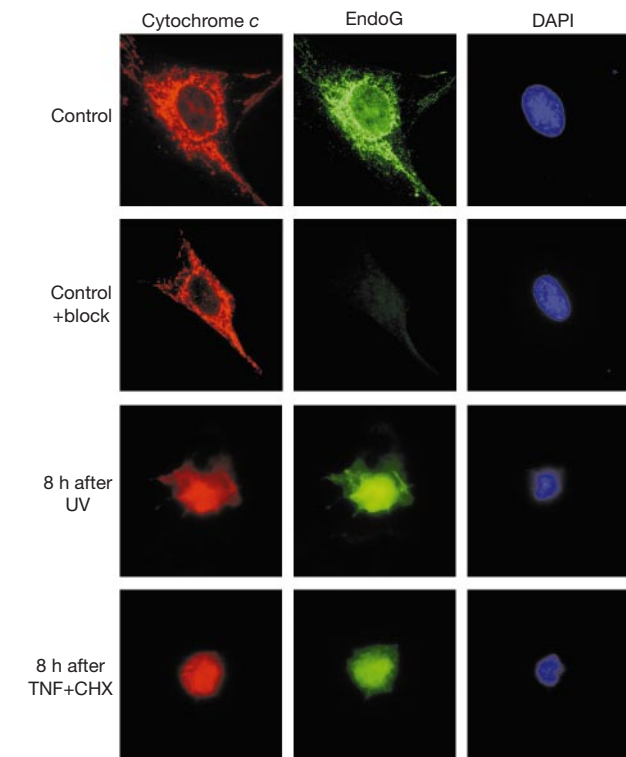


Figure 5 EndoG is released from mitochondria of apoptotic cells during apoptosis. Immunostaining of both endoG and cytochrome *c* (see Methods) was performed on untreated control cells and cells treated with ultraviolet (UV) radiation or TNF/CHX. Control cells were also stained with cytochrome *c* and anti-endoG antibody in the presence of recombinant endoG (block). Cell nuclei were stained with DAPI.

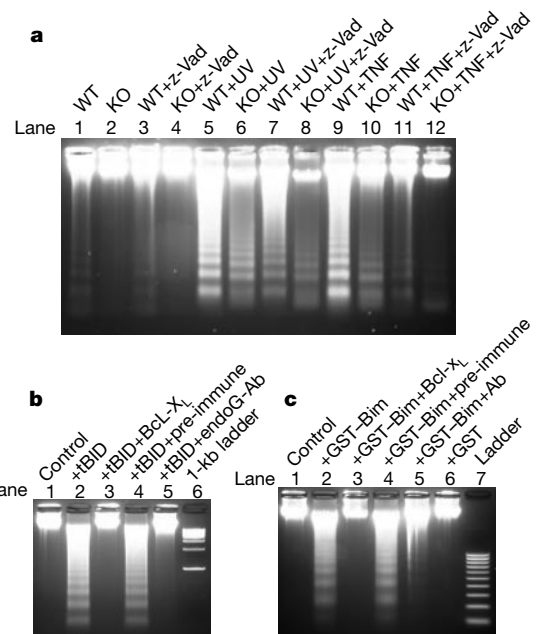


Figure 6 DFF-knockout cells can undergo caspase-independent DNA fragmentation. **a**, Wild-type (WT) and DFF-knockout (KO) MEF cells were treated with ultraviolet radiation or TNF/CHX, respectively (see Methods). z-Vad-fmk (z-Vad) was added to the medium to 50 μ M 4 h before treatment. Cells were collected 24 h after the treatment and processed (see Methods). Aliquots of 20 μ g of the extracted genomic DNA were subjected to 1.2% TBE agarose gel electrophoresis and visualized under ultraviolet light. **b, c**, DFF-knockout cells (2×10^6) were collected by trypsinization and washed three times with cold PBS. The pellet was resuspended in 1 ml of buffer L (20 mM HEPES at pH 7.5, 10 mM KCl, 4 mM MgCl₂, 0.1 mM PMSF and 250 mM sucrose), processed 10 times with a teflon Dounce homogenizer and centrifuged at 12,000g for 15 min at 4 °C. The pellet, containing both nuclei and mitochondria, was resuspended in 400 μ l of buffer L. Aliquots of 53 μ l were then added to the following: buffer L alone (control, **b** and **c**), 400 ng of tBid (**b**) or 600 ng of GST-Bim (**c**); lane 3 was pre-incubated with 800 ng of Bcl-x_L for 15 min at 37 °C, lane 4 included 2 μ l of pre-immune serum, and lane 5 included 2 μ l of anti-endoG antibody (Ab); lane 6 was 600 ng of GST alone. The reactions were incubated at 37 °C for 3 h and the DNA samples were isolated and analysed (see Methods).

efficiently blocked Bim-induced DNA fragmentation while pre-immune serum did not (lanes 3–5).

These experiments revealed endoG as another apoptogenic protein from mitochondria that directly leads to nucleosomal DNA fragmentation. EndoG may serve as an alternative pathway to DFF/CAD to cause genomic DNA fragmentation in a caspase-independent manner.

EndoG has been proposed to participate in mitochondrial replication through the generation of RNA primers required for the initiation of mitochondrial DNA synthesis, an event that takes place in the matrix of the mitochondria¹². However, because tBid induces release of cytochrome *c* from mitochondria without affecting the mitochondrial membrane potential, and because a mitochondrial matrix protein, mtHsp70, was not released by tBid even when most of the cytochrome *c* and endoG had been released, it is likely that most of the endoG co-localizes with cytochrome *c* in the intermembrane space (ref. 16 and Fig. 4). Also consistent with its proposed function in apoptosis, the treatment of DNA with damaging agents such as L-ascorbic acid, peplomycin and cisplatin could enhance the susceptibility of DNA to the nuclease activity of endoG¹⁷. Such a property may be important for endoG to process damaged DNA that may be more resistant to other DNases.

Support has been increasing for the idea that, even though caspases are important in many aspects of programmed cell death, cells can still undergo programmed cell death without caspase activation. The broad-spectrum caspase inhibitor z-Vad-fmk could not prevent cell death induced by the overexpression of Bax to mitochondria even though most of the caspase-mediated apoptotic markers were eliminated¹⁸. The importance of mitochondria was further demonstrated by the identification of several apoptogenic factors. In addition to cytochrome *c*, which activates caspases through its cofactor Apaf-1, mitochondria in cells undergoing apoptosis also release Smac (also known as DIABLO) to counteract the caspase-inhibitory activities of IAPs (inhibitor-of-apoptosis proteins) and AIF (apoptosis-inducing factor) to cause fragmentation of DNA into large sections ($M_r \sim 50K$) and condensation of peripheral nuclear chromatin^{19–21}. Moreover, mice lacking Bak and Bax displayed striking phenotypes, such as the retention of interdigital web in adult animals and accumulation of excess cells in the haematopoietic system, which were not observed in animals with Apaf-1 or caspase-9 knocked out^{23–29}. Examination of the Apaf-1-knockout mice showed necrotic cell death during interdigital web recession without caspase activation³⁰. These results indicate that the mitochondrial damage leading to the release of proteins of the intermembrane space might lead to both caspase-dependent and -independent cell death. EndoG represents such a caspase-independent pathway of cell death initiated from mitochondria. □

Methods

In vitro assay for DNase activity

Mouse liver mitochondria, the full-length recombinant caspase-8 and Bid were prepared as described⁹. A 6- μ l aliquot of purified mouse liver mitochondria was added to 4 μ l of caspase-8 (0.2 mg ml⁻¹) and 1 μ l Bid (1 mg ml⁻¹), or 1 μ l tBid (0.1 mg ml⁻¹) in a final volume of 50 μ l buffer R (20 mM HEPES-KOH at pH 7.5, 220 mM mannitol, 68 mM sucrose, 10 mM KCl, 1.5 mM MgCl₂, 1 mM NaEDTA, 1 mM NaEGTA, 1 mM dithiothreitol (DTT) and 0.1 mM phenylmethyl sulphonyl fluoride (PMSF)). After incubation at 30 °C for 45 min and centrifugation at 12,000g for 15 min, the supernatants from the reactions were used in either the nuclei or plasmid assay. For the nuclear assay, an aliquot of 38 μ l of the supernatant was added to 5 μ l of purified nuclei made as described².

Purification of the supernatant DNase

Liver mitochondria from 30 Swiss Webster mice were prepared and treated with tBid as described above; the treated-mitochondria supernatant (TMS) was used as the starting material for the following purification procedures, with ammonium sulphate precipitation as the first step. Supernatant from 20% precipitation was further precipitated at a final concentration of 60% ammonium sulphate. The pellet from the precipitation was resuspended in 6 ml of buffer A (20 mM HEPES-KOH at pH 7.5, 10 mM KCl, 1.5 mM MgCl₂, 1 mM EDTA, 1 mM EGTA, 1 mM DTT and 0.1 mM PMSF), loaded onto a

Superdex 200 gel-filtration 10/30 column, equilibrated, and eluted with buffer A. Active DNase fractions were pooled and loaded onto a Mono Q 5/5 column equilibrated with buffer A. The flow-through from the column was collected and loaded onto a Mono S 5/5 column and eluted with a linear gradient of 15 ml from buffer A to buffer A containing 250 mM NaCl. An aliquot of 0.5 ml of the active peak fraction was dialysed against buffer A for 2 h and loaded onto the Superdex 200 column equilibrated and eluted with 60 mM NaCl in buffer A. Fractions of 1 ml were collected and assayed for DNase activity.

Protein identification

A 50- μ l aliquot of the active peak fraction from the last purification step was subjected to 15% SDS-PAGE. After staining with silver (Silver-Staining Plus, Bio-Rad), the 30K band that corresponded with the activity was excised from the gel and digested with trypsin. A 1- μ l aliquot from the digest was used for peptide mass fingerprinting as described⁷. Mass values were searched against the NCBI database with the program MS-Fit. The mass values of five peptides (a total of 54 amino acids) matched that of mouse endoG with a standard deviation of <0.34.

Production of recombinant proteins

Full-length human endoG with a six-histidine carboxy-terminal tag was generated as described¹⁹ except that the cell pellet was resuspended and sonicated in buffer A with 0.5% CHAPS. A COOH terminal portion (amino acids 127–286) of mouse endoG with a six-histidine C-terminal tag was cloned into pET-15b and the recombinant protein was purified following the protocol described². Full-length Bcl-x_L complementary DNA was cloned into *NdeI/XhoI* sites of the expression vector pET-15b and the recombinant protein was purified as above. Full-length GST-tagged Bim-EL was cloned into *EcoRI/XhoI* sites of the expression vector pGEX-VP (Novagen) and the recombinant protein was purified using the standard purification protocol for GST-tagged proteins.

Immunodepletion, blotting and immunostaining

Anti-endoG antibodies were generated by immunizing rabbits with the recombinant protein containing the C-terminal portion of endoG prepared as described above. Western blotting and immunodepletion were performed following the protocol described². For immunostaining, DFF45/ICAD-knockout MEF cells were seeded on chamber slides (Nalge Nunc International) in DMEM medium with 10% (v/v) fetal calf serum, and grown overnight in a 37 °C incubator with 5% CO₂. The cells were either irradiated 3.5 cm from an ultraviolet lamp (5 mW cm⁻²; Philips, G36T6L) or incubated with 2 ng ml⁻¹ of TNF- α (Sigma) plus 1 μ g ml⁻¹ of CHX (Sigma). After 8 h, the cells were immunostained as described²⁰.

Received 20 February; accepted 23 May 2001.

- Wyllie, A. H. Glucocorticoid induced thymocyte apoptosis is associated with endogenous endonuclease activation. *Nature* **284**, 555–556 (1980).
- Liu, X. *et al.* DFF, heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**, 175–184 (1997).
- Liu, X. *et al.* DFF45 induces DNA fragmentation and chromatin condensation during apoptosis. *Proc. Natl Acad. Sci. USA* **95**, 8461–8466 (1998).
- Enari, M. *et al.* A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**, 43–50 (1998).
- Zhang, H. *et al.* Resistance to DNA fragmentation and chromatin condensation in mice lacking the DNA fragmentation factor 45. *Proc. Natl Acad. Sci. USA* **95**, 12480–12485 (1998).
- McIlroy, D. *et al.* An auxiliary mode of apoptotic DNA fragmentation provided by phagocytes. *Genes Dev.* **14**, 549–558 (2000).
- Zhang, J. H. & Xu, M. DNA fragmentation in apoptosis. *Cell Res.* **10**, 205–211 (2000).
- Li, H. *et al.* Cleavage of Bid by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* **94**, 615–624 (1998).
- Luo, X. *et al.* Bid, a Bcl2 interacting protein, mediates cytochrome *c* release from mitochondria in response to activation of cell surface death receptor. *Cell* **94**, 481–490 (1998).
- Gross, A. *et al.* Caspase cleaved BID targets mitochondria and is required for cytochrome *c* release, while BCL-XL prevents this release but not tumor necrosis factor-R1/Fas death. *J. Biol. Chem.* **274**, 1156–1163 (1999).
- Tiranti, V. *et al.* Chromosomal localization of mitochondrial transcription factor A (TCF6), single-stranded DNA-binding protein (SSBP), and endonuclease G (ENDOG), three human housekeeping genes involved in mitochondrial biogenesis. *Genomics* **25**, 559–564 (1995).
- Cote, J. & Ruiz-Carrillo, A. Primers for mitochondrial DNA replication generated by endonuclease G. *Science* **261**, 765–769 (1993).
- Krammer, P. H. CD95's deadly mission in the immune system. *Nature* **407**, 789–795 (2000).
- Kluck, R. M. *et al.* The release of cytochrome *c* from mitochondria: a primary site for Bcl-2 regulation of apoptosis. *Science* **275**, 1132–1136 (1997).
- Puthalakath, H. *et al.* The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex. *Mol. Cell* **3**, 287–296 (1999).
- Von Ahsen, O. *et al.* Preservation of mitochondrial structure and function after Bid- or Bax-mediated cytochrome *c* release. *J. Cell Biol.* **150**, 1027–1036 (2000).
- Ikeda, S. & Ozaki, K. Action of mitochondrial endonuclease G on DNA damaged by L-ascorbic acid, peplomycin, and cis-Diamminedichloroplatinum (II). *Biochem. Biophys. Res. Commun.* **235**, 291–294 (1997).
- Xiang, J., Chao, D. T. & Korsmeyer, S. J. BAX-induced cell death may not require interleukin 1 beta-converting enzyme-like proteases. *Proc. Natl Acad. Sci. USA* **93**, 14559–14563 (1996).
- Zou, H. *et al.* An APAF-1 cytochrome C multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
- Du, C. *et al.* Smac, a mitochondrial protein that promotes cytochrome *c*-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).

21. Verhagen, A. *et al.* Identification of DIABLO, a mammalian protein that promotes apoptosis by binding to and antagonizing IAP proteins. *Cell* **102**, 43–53 (2000).
22. Susin, S. A. *et al.* Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397**, 441–446 (1999).
23. Lindsten, T. *et al.* The combined functions of proapoptotic Bcl-2 family members Bak and Bax are essential for normal development of multiple tissues. *Mol. Cell* **6**, 1389–1399 (2000).
24. Kuida, K. *et al.* Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* **384**, 368–372 (1996).
25. Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways *in vivo*. *Cell* **94**, 339–352 (1998).
26. Kuida, K. *et al.* Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* **94**, 325–337 (1998).
27. Cecconi, F. *et al.* Apaf1 (CED-4 homolog) regulates programmed cell death in mammalian development. *Cell* **94**, 727–737 (1998).
28. Yoshida, H. *et al.* Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* **94**, 739–750 (1998).
29. Honarpour, N. *et al.* Adult Apaf-1-deficient mice exhibit male infertility. *Dev. Biol.* **218**, 248–258 (2000).
30. Chautan, M. *et al.* Interdigital cell death can occur through a necrotic and caspase-independent pathway. *Curr. Biol.* **9**, 967–970 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. M. Peters, M. Lutter and M. Fang for their assistance in mitochondrion purification and other techniques; Y. Li and R. Harold for technical support; J. Zhang and M. Xu for providing DFF45-knockout MEF cells. We also thank M. Lutter and X. Jiang for recombinant Bcl-x_L. X.L. is supported by the Leukemia Society of America; X.W. is supported by grants from NIH and the Welch Foundation.

Correspondence and requests for materials should be addressed to X.W. (e-mail: xwang@biochem.swmed.edu).

corrections

Timing of the Last Glacial Maximum from observed sea-level minima

Yusuke Yokoyama, Kurt Lambeck, Patrick De Deckker, Paul Johnston & I. Keith Fifield

Nature **406**, 713–716 (2000).

In this Letter, a statement on page 715 that the eustatic sea level $\Delta\zeta_{\text{eust}}(t)$ differs from the ice-volume-equivalent sea level $\Delta\zeta_{\text{e}}(t)$ is erroneous. We thank W. R. Peltier for drawing this error in interpretation to our attention. The two definitions were initially introduced to distinguish changes in sea level that were the result of ice mass being added to the oceans from other changes caused by thermal expansion, for example. The erroneous statement was the result of a programming error that was introduced into the program in 1998 when it was modified to deal with shelf ice more accurately and to introduce the time dependence of the coastlines in the evaluation of the equivalent sea-level change through equation (3). This evaluation requires a knowledge of the location of the shoreline at each time step, of the grounding line of the ice and of the part of the shelf ice that floats or grounds as a result of sea level change and shelf-ice thickness change. The error was introduced at the end of each iteration of the sea-level equation, when the value of sea level

change was integrated over the ocean and the resulting volume was compared with the change in ice volume for the corresponding interval. The eustatic sea level at this stage was defined as the global average of sea-level change but, because we allow sea level to be non-zero on land so as to be able to compute the tilting of lakes or the changing elevations of tree lines, the averaging should have been restricted to the ocean. Fortunately, the error does not enter into any other part of the model predictions because the estimate of global sea level rise is based on the ice volume change in equation (3) throughout the core program. Thus calculations of the isostatic correction $\Delta\zeta_{\text{i}}(j,t)$ in equation (1) and of the ice-volume-equivalent sea level $\Delta\zeta_{\text{e}}(t)$ are correct, as are the results illustrated in Fig. 2. The matter arises in the two sentences in parentheses of the penultimate paragraph of the main text (page 715), in a discussion of the difference between our Barbados results and those of Peltier. The first sentence in parentheses remains correct, but our attempt to explain it in the second sentence is not. The cause for the disagreement must be sought elsewhere, possibly in the different ice and/or earth models used. □

The role of interleukin-1 polymorphisms in the pathogenesis of gastric cancer

Emad M. El-Omar, Mary Carrington, Wong-Ho Chow, Kenneth E. L. McColl, Jay H. Bream, Howard A. Young, Jesus Herrera, Jolanta Lissowska, Chiu-Chin Yuan, Nathaniel Rothman, George Lanyon, Maureen Martin, Joseph F. Fraumeni Jr & Charles S. Rabkin

Nature **404**, 398–402 (2000).

We reported the association of interleukin-1 gene cluster polymorphisms suspected of enhancing production of interleukin-1 β with an increased risk both of hypochlorhydria induced by *Helicobacter pylori* and of gastric cancer. We were subsequently alerted to an error in our genotyping by a report¹ that the *IL-1B*-511T variant is in positive linkage disequilibrium with the *IL-1B*-31C allele, rather than with the *IL-1B*-31T allele as we had stated. We have traced this discrepancy to incorrect labelling of the sequence data from the wild-type and variant controls for the *IL-1B*-31 assays. Consequently, the designations *IL-1B*-31T and *IL-1B*-31C were reversed in the text and Tables 1–3. We have now verified the proper identifications by re-sequencing our controls. In addition, we have confirmed the more common (*IL-1B*-31T/*IL-1B*-511C) and less common (*IL-1B*-31C/*IL-1B*-511T) haplotypes by forward and reverse sequencing five homozygous samples (two wild-type and three variant) for both of these loci. Our data and analysis remain otherwise unchanged. The electrophoretic mobility shift assays were unaffected by this error and hence the greater DNA binding of the *IL-1B*-31T-bearing oligonucleotide represents the wild-type *IL-1B* promoter. □

1. Hamajima, N. *et al.* Interleukin-1 polymorphisms, lifestyle factors, and *Helicobacter pylori* infection. *Jpn J. Cancer Res.* **92**, 383–389 (2001).

The heat is on

Thermocycling and PCR in general are the themes for this week's selection.

ExpressChip DNA microarrays

a1-biotech/Mergen www.mergen-ltd.com

Focal objectives

Each ExpressChip DNA microarray kit contains two identical glass microarray slides, reagents for hybridization and detection, hybridization chambers, Cy3-labelled reagents and an instruction manual. Slides are pre-spotted with sequences functionally important for cell metabolism and/or disease development, and include human housekeeping and bacterial genes as positive, negative and normalization controls. Kits cover known genes of human, mouse or rat and are mapped to the UniGene/ GenBank database. Focused gene groups include those for cell cycle proteins, growth factors, cytokines/receptors and disease-related genes. ExpressChip is produced jointly by a1-biotech and Mergen.

PCR Master Mix

Promega www.promega.com

Just add water

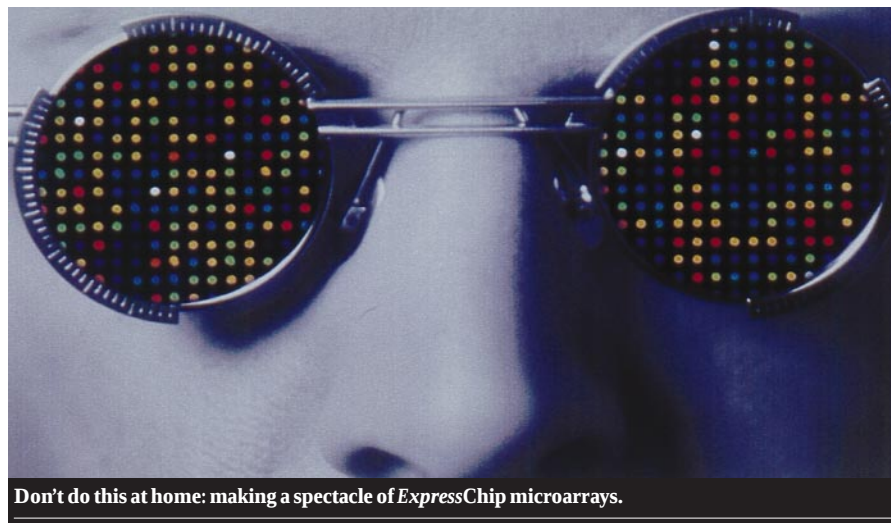
This convenient pre-mixed formulation with Taq DNA polymerase is a 2 × solution containing all PCR reagents and requiring only the addition of primers, template and water. It is sensitive to as little as 1.5 copies of a target template, and scalable for 10, 25 and 50 µl reactions. Storage is at 4°C for up to three months.

Touchgene thermal cycler

Techne www.techneuk.co.uk

Fingertip temperature control

A large (115 × 90 mm), touch-sensitive interface and a gradient range of 30°C are the major features of this thermal cycler. Four independent temperature sensors control eight Peltier units. The graphical display shows the sample temperature profile in real time while the program is running, including upper and lower limits



Don't do this at home: making a spectacle of ExpressChip microarrays.

of the gradient. A calculator function displays the temperature of each column and active ramping and fourfold control ensure that each sample is at the target temperature for the same hold time. Maximum ramp rate is more than 3°C per second.

HTP TOPO kit

Invitrogen www.invitrogen.com

Clever cloning

TOPO cloning technology, which provides five-minute, bench-top ligation of PCR products with ≥95% recombinants, can also be used for high-throughput cloning. These kits contain 500 reactions of TOPO vector in a single tube. The reactions can be set up in multiwell plates, incubated for five minutes and then transformed into competent *E. coli*. The protocol also supports robotic procedures. Four different vectors are available, with TOP10 *E. coli* supplied either in bulk or in 96-well plates.

MasterAmp

Epicentre www.epicentre.com

High-sensitivity and high-fidelity RT-PCR kits

The MasterAmp RT-PCR kit provides high-temperature cDNA synthesis and amplification using a single-tube, single-buffer system to minimize chances of contamination. For applications needing high fidelity, the MasterAmp high-fidelity RT-PCR kit is available. First-strand synthesis is catalysed by Moloney murine leukaemia virus reverse transcriptase (MMLV-RT) and the resulting cDNA product amplified using MasterAmp TAQurate DNA polymerase blend. Both kits permit amplification of templates considered difficult, such as those with a high G + C content or extensive secondary structure. Kits are supplied for 100 reactions.

ProofStart

Qiagen www.qiagen.com

Proven protection for primers

Proofreading enzymes utilize a 3'-to-5' exonuclease activity to remove incorrectly incorporated bases. This activity can also degrade primers during setup and start of PCR, causing non-specific priming, smearing, or failure to amplify any product at all. ProofStart DNA polymerase is chemically modified to inactivate both the polymerase and the exonuclease activity. Primers remain intact and mispriming cannot occur during setup. A 'hot-start' feature allows PCR to start with high proofreading activity.

Micromax slide scanning

PerkinElmer www.perkinelmer.com

Scans via the web

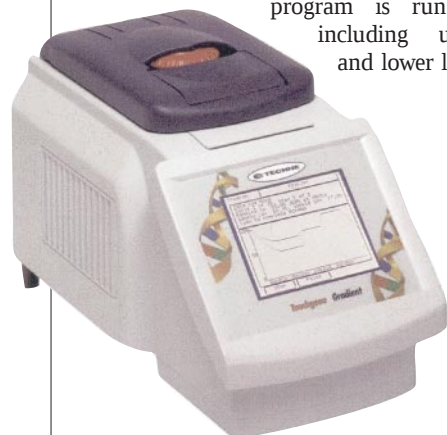
Data from the Micromax microarray scanning service can now be obtained via the Internet, enabling researchers to use the glass slide-based microarray systems for high-throughput, differential gene expression analysis without investing in scanning equipment. Cyanine-3 and cyanine-5 channels are scanned sequentially and results saved to both JPG and 16-bit TIFF files. Scans are performed at optimized laser power and photomultiplier settings. Results are delivered to the user's secure, password-protected site, and can be shared with assigned colleagues.

Rat Foundation 1

Incyte Genomic www.incyte.com

Rattus norvegicus microarray

This broad-based microarray is designed to assay the expression of genes from the Sprague-Dawley rat. It is constructed with



Techne: offering touch-screen control

new on the market

non-redundant genes that provide the data needed to monitor the expression patterns of over 9,000 rat genes in a single experiment.

Falcon

BD Biosciences www.bdbiosciences.com
PCR labware

The Falcon line of labware offers a range of products for PCR, from 0.5-ml tubes to 384-well microplates, along with a variety of accessories for sealing PCR plates and for low-stress handling of PCR vessels. All vessels are tested for DNase, RNase and human genomic DNA.

5' EndTag

Vector Laboratories www.vectorlabs.com
Nucleic acid labelling system

With this system labels can be attached only at the 5' end, leaving the 3' end available for polymerization. The position of the label does not interfere with hybridization or nucleic acid binding. The system is designed for labelling PCR and sequencing primers and a variety of labels including biotin, fluorescein, Texas Red, DNP, fucose or coumarin 151 can be attached.

McSNP

Thermo Hybaid www.hybaid.com
Genotyping is a snip with this system

The McSNP (melting curve analysis of single nucleotide polymorphisms) system combines the PCR-RFLP approach with fluorescence-based, automatic, real-time SNP scoring. It offers auto-scoring of both homozygotes and heterozygotes, simple assay optimization and automated PCR-RFLP analysis. Up to 20,000 SNPs can be run per week, and the company claims accuracy and reproducibility of 99% and 100%, respectively.

Mastercycler 384

Eppendorf www.eppendorf.com
Wall-to-wall thermocycling

Suitable for use in laboratories with a high throughput, this 384 thermal block instrument



Falcon labware, available at bdbiosciences.com.

is designed so that the walls of all 384 tubes are in close contact with the metal walls of the thermocycler, accelerating thermal transfer to the sample and maximizing the homogeneity and reliability of the results. A heated lid prevents condensation in the upper areas of the tube. A gradient PCR option can also be used.

I-SAGE

Invitrogen www.invitrogen.com
Sage advice on gene expression

I-SAGE is a pre-assembled, pre-tested kit that can perform quantification of every gene expressed in a cell, whether known or novel. Library construction is convenient since there is no need to search for, order and test numerous reagents. The kit can be used to find low-abundance transcripts, detect up- or downregulated genes, measure a compound's effect on expression and identify novel genes.

PCR Prep Station

Labcaire www.labcaire.co.uk
Cabinet shuffle

Designed to protect both the operator and the samples from contamination, this new cabinet features a mini vertical laminar flow allowing users to reach samples easily while being protected from potential hazards. A recirculating airflow through sealed HEPA filters offers fur-

ther protection. The work surface is bathed continuously in clean air and an automatic timed UV light facility can be activated with the safety door in position to destroy potential contamination and eliminate the risk of carry-over during sample preparation. The unit is compact enough for the work-bench, is quiet in operation and requires no external ductwork.

Fast Start

Roche Molecular Biochemicals
www.biochem.roche.com/pcr

Get rich, quick

Fast Start is designed to maximize specificity in hot-start reactions, eliminate wax barriers, beads, hot-start antibodies, manual hot-start and the need to set up PCR reactions on ice. By combining the enzyme with the company's optimized buffer system and GC-RICH resolution solution, problematic templates such as GC-rich sequences, repeats, and those with high secondary structure can be amplified. The product is suitable for routine, multiplex, nested, high-throughput or complex genomic PCR on fragments up to 3 kb.

Autopure LS

Gentra Systems www.gentra.com
Get it large

Autopure LS is a nucleic acid purification instrument for the extraction of genomic DNA from a wide range of large samples, such as 1 to 10 ml of blood. It can produce DNA yields of up to 500 µg and process 96 samples in just 8 hours. Autopure LS carries an on-board, refrigerated centrifuge that loads and unloads automatically. The system uses Puregene chemistry with sophisticated robotics to produce high yields of high molecular weight DNA (100–200 kb) from a wide range of large sample types. Puregene effectively removes PCR inhibitors and other contaminants, making the DNA suitable for a variety of downstream applications and Southern blotting. Tracking and validation data are reported to a PC workstation.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENTS

**DIVERSE
SMALL
MOLECULES
READY FOR
SCREENING**

**10K–50K
Compounds
in Plates**

• Next Day Delivery*
• Competitively Priced
(*See our Standard Conditions)

Chembridge Corporation • 800.954.6943
950-49-3400 • 950-49-3400 ext. • E-mail: chem@chembridge.com

www.chembridge.com

**Polyclonal Antibodies
& Custom Peptides**

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen
Invitrogen Corporation

U.S. or Canada 800-855-4368
FAX 256-536-9016

If your work requires reagents for:

- ~ proteasome ~
- ~ ubiquitin ~
- ~ COP9-signalosome ~

research

we have the most comprehensive range of antibodies ~ biochemicals ~ kits

www.proteasome.com

AFFINITY Research Products Ltd.
Marnhead Castle, Exeter, EX8 8HD, U.K.
T: +44(0) 1626 891010 ~ F: +44(0) 1626 891090
~ Email: info@proteasome.com ~